

# What is the Real Question?

## An Empirical-Ontological Approach to the Interpretative Analysis of Archival Reference Questions

Dissertation

zur Erlangung des akademischen Grades  
Doctor philosophiae (Dr. phil.)

im Fach Bibliotheks- und Informationswissenschaft  
eingereicht an der  
Philosophischen Fakultät I  
Humboldt-Universität zu Berlin

von  
Steffen Hennicke

Präsident der Humboldt-Universität zu Berlin:  
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Philosophischen Fakultät I:  
Prof. Michael Seadle, Ph.D.

Gutachter

1. Prof. Vivien Petras, Ph.D.
2. Dr. Martin Doerr
3. Prof. Allen Renear, Ph.D.

Datum der Einreichung: 08.09.2015

Datum der Disputation: 26.01.2016

## Abstract

Research on the information behaviour of archive users has repeatedly shown that access to archival knowledge bases such as via finding aids is hindered by a certain deficiency in knowledge about users and their information needs. The objective of this study is to increase our empirical understanding of the information needs of archive users in order to facilitate access to archival information systems. The primary research question explores the possibility of creating an ontology that can reasonably represent archival inquiries and their probable interpretations as formal queries to the model of the archival target world that would generate an adequate response. For this purpose, the study analyzes archival information needs as expressed in written natural language questions submitted to the German Federal Archives (*Bundesarchiv*) and the National Archives of Norway (*Arkivverket*). The primary results point to a methodological approach that advocates an interpretative analysis of written natural language inquiries and, as the result of its practical and successful application, the *Archival Knowledge Model* (AKM). This constitutes an ontological representation of the subject matter typically appearing as the interest of archival inquiries. The study thus shows that written natural language questions submitted to archives exhibit dominant fundamental ontological patterns, which can be formalized to a relatively small set of entities utilizing the CIDOC CRM and compatible extensions wherever its semantics are found to be inadequate. The AKM formally and explicitly represents the information needs of archive users in the form of an ontology, thus providing a constitutive model for the evaluation of existing archive information systems and data schemas and may inform the development of new cataloguing rules or pattern-based query applications.

## Zusammenfassung

Die Forschung zum Informationsverhalten in Archiven hat wiederholt gezeigt, dass der Zugang zu Archivbeständen durch einen grundlegenden Wissensmangel über die Informationsbedürfnisse von Archivnutzern eingeschränkt wird. Das Ziel der vorliegenden Studie ist die Vertiefung des empirischen Verständnisses der spezifischen Informationsbedürfnisse von Archivnutzern und damit die Leistung eines grundlegenden Beitrags zur Verbesserung von archivischen Informationssystemen. Dazu untersucht die primäre Forschungsfrage die Möglichkeit der Erarbeitung einer Ontologie, die in angemessener Weise das Erkenntnisinteresse von Nutzern an Archive repräsentiert, und auf formale Anfragen, adäquate Antworten erzeugt. Zu diesem Zweck analysiert die Studie schriftliche Nutzeranfragen an das Bundesarchiv und das Nationalarchiv von Norwegen (*Arkivverket*). Die primären Forschungsergebnisse umfassen einen methodischen Ansatz, der eine interpretative Analyse schriftlich formulierter Fragen an Archive vertritt und, als Ergebnis seiner erfolgreichen Anwendung, das *Archival Knowledge Model* (AKM) erarbeitet. Das AKM ist eine ontologische Repräsentation des Gegenstandsbereichs, der typischerweise als das Erkenntnisinteresse von Fragen an Archiven erscheint. Die Studie zeigt, dass derartige Anfragen dominante, grundlegende ontologische Muster aufweisen, die auf einen relativ kleinen Satz ontologischer Entitäten formalisiert werden können. Das AKM repräsentiert damit formal und explizit die Informationsbedürfnisse von Archivnutzern in Form einer Ontologie. Diese stellt ein konstitutives Modell für die Evaluierung bestehender Archivinformationssysteme und Datenschemata zur Verfügung und kann die Entwicklung neuer Katalogisierungsregeln oder musterbasierter Abfrageanwendungen unterstützen.

# Table of Contents

List of Figures . . . . .	v
List of Tables . . . . .	viii
List of Abbreviations . . . . .	ix
<b>I Introduction</b>	<b>1</b>
<b>II Related Research</b>	<b>8</b>
1 Information Behaviour . . . . .	8
2 Archival Information Behaviour . . . . .	10
3 Archival Information Needs . . . . .	13
4 Archival Implementations of EAD . . . . .	15
5 Summary: Global Digital Access . . . . .	18
<b>III Empirical Data</b>	<b>21</b>
1 Reference Questions . . . . .	21
2 Data Collection . . . . .	23
2.1 The German Federal Archives . . . . .	24
2.2 The National Archives of Norway . . . . .	26
3 Single Inquiries . . . . .	30
4 Summary: Sample Data . . . . .	32
<b>IV Methodological Approach</b>	<b>34</b>
1 Interpretative Analysis . . . . .	35
1.1 Linguistic Analysis . . . . .	36
1.1.1 Types of Questions . . . . .	36
1.1.2 Information Entities . . . . .	42
1.2 Ontological Analysis . . . . .	43
1.2.1 Referenced Entities . . . . .	45

1.2.2	Patterns . . . . .	47
1.3	Summary . . . . .	49
<b>2</b>	<b>Epistemological Framework . . . . .</b>	<b>50</b>
2.1	Historical Domain . . . . .	51
2.2	Archival Domain . . . . .	57
2.3	Summary . . . . .	67
<b>3</b>	<b>Ontological Modelling . . . . .</b>	<b>68</b>
3.1	Modelling Methodology . . . . .	69
3.2	Classes and Properties . . . . .	74
3.3	Notation . . . . .	81
3.4	Summary . . . . .	85
<b>4</b>	<b>Summary: Knowledge Engineering . . . . .</b>	<b>86</b>
<b>V</b>	<b>Analysis and Results . . . . .</b>	<b>91</b>
<b>1</b>	<b>Interpretative Analysis . . . . .</b>	<b>91</b>
1.1	Types of Questions . . . . .	91
1.2	Referenced Entities . . . . .	95
1.2.1	Given Entities . . . . .	95
1.2.2	Wanted Entities . . . . .	102
1.3	Interest of Inquiries . . . . .	111
1.4	Ontological Core Framework . . . . .	116
1.5	Provenance and Aboutness . . . . .	121
1.6	Fundamental Phenomena . . . . .	124
1.7	Summary . . . . .	127
<b>2</b>	<b>General Patterns . . . . .</b>	<b>130</b>
2.1	Provenance Context . . . . .	130
2.1.1	Provenance . . . . .	131
2.1.2	Documentation . . . . .	143
2.1.3	Self-Documentation . . . . .	154
2.1.4	Correspondence . . . . .	162
2.2	Aboutness Context . . . . .	170
2.2.1	Aboutness . . . . .	171
2.2.2	Events . . . . .	175
2.2.3	Plans . . . . .	184
2.2.4	Actors . . . . .	202

2.2.5	Things	210
2.2.6	Documents	215
2.3	Summary	222
<b>3</b>	<b>Summary: Archival Knowledge Model</b>	<b>230</b>
<b>VI</b>	<b>Application</b>	<b>235</b>
<b>1</b>	<b>Encoded Archival Description</b>	<b>236</b>
<b>2</b>	<b>Archival Data and the AKM</b>	<b>237</b>
<b>3</b>	<b>Summary: Querying and Implementation</b>	<b>245</b>
<b>VII</b>	<b>Conclusion</b>	<b>249</b>
	<b>Appendix: Definition of the AKM</b>	<b>257</b>
	Native AKM Classes	257
	Native CRM Classes	260
	Native AKM Properties	265
	Native CRM Properties	268
	<b>References</b>	<b>278</b>

## List of Figures

1	Email sent to the National Archives of Norway. . . . .	30
2	Letter attached to an email containing an inquiry with one question and associated contextual information. . . . .	31
3	Model of the levels of archival arrangement according to ISAD(G) (International Council of Archives (2000, 36)). . . . .	63
4	The top-level classes of the CRM, taken from Doerr (2003, 85). . . . .	74
5	The event-centric approach of the CRM, taken from Doerr (2003, 87). . . . .	76
6	Basic notation for general patterns. . . . .	83
7	Additional notational elements for query and archival patterns. . . . .	85
8	The AKM as a mediator between the historical and the archival domain. . . . .	88
9	Primary categories of question types detailing non-discovery questions. . . . .	92
10	Primary categories of question types detailing discovery questions. . . . .	93
11	Substantial types of questions per sample. . . . .	94
12	Groups of given entities per total sample (n=1656). . . . .	99
13	Detailed listing of the three largest groups of given entities per total sample (percentages given per primary category). . . . .	100
14	Groups of given entities per individual sample. . . . .	101
15	Primary categories of wanted entities. . . . .	106
16	Primary categories of wanted entities per sample. . . . .	107
17	Details of the wanted entities group Actors. . . . .	108
18	Details of the wanted entities group Activities. . . . .	109
19	Primary categories of wanted entities per primary categories of type of questions. . . . .	110
20	The level of interest of inquiries. . . . .	115
21	Ontological core framework of the <i>Archival Knowledge Model</i> . . . . .	117
22	Groups or persons carrying out an activity. . . . .	125
23	Actors carrying out an activity. . . . .	126
24	General <i>Provenance</i> pattern. . . . .	132
25	<i>Provenance</i> pattern: Types of information objects created. . . . .	133
26	<i>Provenance</i> pattern: Time and place for creation and production activities. . . . .	137
27	Q003-06-02 <sub>barch</sub> : Any personal materials by a particular person. . . . .	139
28	Q059-01-04 <sub>barch</sub> : All things kept by a group. . . . .	141
29	General <i>Documentation</i> pattern. . . . .	144
30	Q002-03-01 <sub>barch</sub> : Surveillance as an example of the observation of others. . . . .	148
31	Q050-01-04 <sub>barch</sub> : Keeping records on a type of group. . . . .	150
32	Q011-01-04 <sub>barch</sub> : Creating a dossier about anti-fascist activists. . . . .	151
33	Q054-04-01 <sub>barch</sub> : Documents about the denazification of a person. . . . .	152
34	General <i>Self-Documentation</i> pattern. . . . .	155
35	Q002-01-04 <sub>barch</sub> : Minutes of the Reichstag. . . . .	157

36	Q002-05-02 <sub>barch</sub> : Minutes of committee meetings. . . . .	158
37	Q005-08-02 <sub>barch</sub> : Report on a trip to France. . . . .	160
38	General <i>Correspondence</i> pattern. . . . .	163
39	Q003-01-02a <sub>barch</sub> : Things created by an actor and <i>designated</i> for another actor. . .	165
40	Q003-01-02b <sub>barch</sub> : Things created by an actor and <i>designated</i> for another actor. . .	165
41	Q009-01-11a <sub>barch</sub> : Documents as the products of <i>contacts</i> with authorities. . . . .	167
42	Q009-01-11b <sub>barch</sub> : Documents as the products of <i>contacts</i> with authorities. . . . .	167
43	Q039-01-03 <sub>barch</sub> : Petitions as special acts of direct communication. . . . .	168
44	Core of the <i>Aboutness</i> pattern. . . . .	171
45	General <i>Aboutness</i> pattern. . . . .	172
46	Q002-01-09 <sub>barch</sub> : Reports about anything. . . . .	174
47	General <i>Events</i> pattern. . . . .	176
48	Q031-01-03a <sub>barch</sub> : Photographs of a particular activity. . . . .	179
49	Q031-01-03b <sub>barch</sub> : Photographs of a particular activity as the result of a docu- mentation activity. . . . .	179
50	Q023-12-02a <sub>barch</sub> : Documents about a particular activity. . . . .	181
51	Q023-12-02b <sub>barch</sub> : Documents about a particular activity. . . . .	181
52	Q059-01-12 <sub>barch</sub> : Documents about a named activity. . . . .	182
53	General <i>Plans</i> pattern . . . . .	185
54	General <i>Plans</i> pattern: Planning, plans, and expressions of will. . . . .	188
55	Q005-08-01 <sub>barch</sub> : Documents about planning activities for a specific event. . . . .	191
56	Q005-08-03 <sub>barch</sub> : Documents attesting discussion or approval of a plan. . . . .	193
57	Q036-01-02 <sub>barch</sub> : Aims of a group. . . . .	194
58	Q036-01-03 <sub>barch</sub> : Course of action of a group. . . . .	195
59	Q036-01-03b <sub>barch</sub> : Course of action of a group. . . . .	196
60	Q036-01-03c <sub>barch</sub> : Course of action of a group. . . . .	197
61	Q036-01-03a <sub>barch</sub> : Course of action of a group. . . . .	198
62	Q014-04-06 <sub>barch</sub> : Decisions. . . . .	199
63	Q002-03-03 <sub>barch</sub> : Factual questions about the name of an actor entrusted with a specific mandate. . . . .	200
64	General <i>Actors</i> pattern. . . . .	203
65	Q059-05-01 <sub>barch</sub> : Documents about a type of group. . . . .	206
66	Q010-01-01a <sub>barch</sub> : Any things about a particular person. . . . .	207
67	Q010-01-01b <sub>barch</sub> : Any things about the activities of a particular person. . . . .	207
68	Q003-05-01a <sub>barch</sub> : Membership of a person in a political party. . . . .	208
69	Q003-05-01b <sub>barch</sub> : Party membership card of a person. . . . .	209
70	General pattern <i>Things</i> . . . . .	210
71	Q013-02-03 <sub>nan</sub> : The fate of submarines. . . . .	213
72	Q047-01-01 <sub>nan</sub> : Information about two villas. . . . .	214

73	General <i>Documents</i> pattern. . . . .	217
74	Q039-07-03 <sub>barch</sub> : Transfer of custody between two institutions. . . . .	219
75	Q052-01-02 <sub>barch</sub> : The extent of files. . . . .	220
76	Q003-05-09 <sub>barch</sub> : Containment of documents. . . . .	221
77	General patterns belonging to the provenance context. . . . .	223
78	General patterns belonging to the aboutness context per sample. . . . .	224
79	General patterns belonging to the aboutness context per major type of question. . . . .	225
80	Combinations of general patterns (percentages per general pattern from the provenance context). . . . .	226
81	General patterns of the provenance context with primary entities of interest. . . . .	228
82	General patterns of the aboutness context with primary entities of interest. . . . .	229
83	XML-Snippet from EAD file “Roter Koffer” . . . . .	238
84	The information from the <unittitle> represented explicitly (EAD file “Roter Koffer”). . . . .	239
85	XML-Snippet from EAD file “labw-f-1-5717” . . . . .	240
86	The information from the <unittitle> represented explicitly (EAD file “labw-f-1-5717”). . . . .	241
87	XML-snippet from EAD file “labw-f-1-603494” . . . . .	242
88	The information from the <unittitle> represented explicitly (EAD file “labw-f-1-603494”). . . . .	243
89	XML-Snippet from EAD file “labw-f-1-5721” . . . . .	244
90	The information from the <unittitle> represented explicitly (EAD file “labw-f-1-5721”). . . . .	245



## List of Tables

1	Given Entities . . . . .	95
2	Wanted Entities . . . . .	102
3	Provenance context and aboutness context general patterns. . . . .	122
4	General patterns of the provenance context. . . . .	131
5	Occurrences of the general pattern <i>Provenance</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	142
6	Occurrences of the general pattern <i>Documentation</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	153
7	Occurrences of the general pattern <i>Self-Documentation</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	161
8	Occurrences of the general pattern <i>Correspondence</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	169
9	General patterns belonging to the provenance context. . . . .	170
10	Occurrences of the general pattern <i>Aboutness</i> . . . . .	175
11	Occurrences of the general pattern <i>Events</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	183
12	Occurrences of the general pattern <i>Plans</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	201
13	Occurrences of the general pattern <i>Actors</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	209
14	Occurrences of the general pattern <i>Things</i> . Note that percentages for the primary entities of interest are per current general pattern. . . . .	214
15	Occurrences of the general pattern <i>Aboutness</i> . . . . .	222
16	The general patterns according to their primary epistemic focus. . . . .	251
17	Native AKM Classes . . . . .	257
18	Native CRM Classes . . . . .	260
19	Native AKM Properties . . . . .	265
20	Native CRM Properties . . . . .	268

## List of Abbreviations

AKM	Archival Knowledge Model
BArch	German Federal Archives ( <i>Bundesarchiv</i> )
CRM	International Committee for Documentation Conceptual Reference Model (CIDOC CRM)
EAD	Encoded Archival Description
FRBR	Functional Requirements for Bibliographic Records
ISAD(G)	General International Standard Archival Description
LaBW	Federal State Archives of Baden-Wuerttemberg ( <i>Landesarchiv Baden-Württemberg</i> )
NAN	National Archives of Norway ( <i>Arkivverket</i> )
OEP	Electronic Public Records ( <i>Offentlig Elektronisk Postjournal</i> )
SAPMO	Foundation Archives of Parties and Mass Organisations of the GDR ( <i>Stiftung Archiv der Parteien und Massenorganisationen der DDR im Bundesarchiv</i> )

## Chapter I

# Introduction

### Motivation

Archives are the cultural memory repositories of modern societies and a reservoir of vast quantities of primary evidence of our past. Research on the past will consult the archive in order to discover primary sources as evidence and to find answers to historical questions. Enabling efficient and successful access for their users in terms of discovering and retrieving as well as the use of archival materials from their holdings is a pivotal function for archives (Finch, 1994, 1-2).

The main means for discovering and accessing relevant primary sources in an archive are so-called *finding aids* and *holding guides* supported by the expertise of the archivist. These *archival aids* are descriptive tools which give an account of the context of creation and provenance of aggregations of archival materials which have been created or collected by one or more actors over a certain time frame in the past. The textual descriptions in archival aids allow researchers and archivists alike to assess the *information potential* (Menne-Haritz, 2001) of archival holdings for their research interests, and to locate and discover potentially relevant sets of archival materials.

The information access enabled by archival aids can be characterized as indirect since archival descriptions typically do not describe the content of single items but of large aggregations and their context within an archival holding. In contrast, library catalogues provide subject access to single items such as monographs. This more direct approach assumes that the information needs of the user either directly correspond to the subject or are adequately generalized by it. However, the character of archival materials as primary sources of evidence and less as informational sources for particular answers, together with their ever-growing quantities, render subject access difficult in the context of archives. Furthermore, the information needs of archive users are not necessarily adequately generalized, nor are they always met by plain subjects.

Even though the conceptualization of archival aids is based on elaborated and historically grown archival principles and models, their design is less informed by explicit and broad knowledge about the information needs of archive users (Cox, 2008). Archivists “have traditionally decided unilaterally what is good for patrons” (Dryden, 2005, 87) and their past success within a relatively closed archival world did not necessitate effective and systematic user studies. Furthermore, the expertise of archivists as well as of archival theory evolves around “the record, not its secondary use nor the various types of researchers who visit archives seeking information” (Duff, 2002, 332). The advent of the digital age, however, has made this deficit apparent, especially since it had been widely, and for that matter, falsely, assumed that the print paradigm of archival aids would translate into the digital age (Anderson, 2004, 83).

The digital age and new Web technologies have yielded various external challenges for archives; for example, the pursuit of extended outreach and the need to facilitate access for

new and more disparate and independent user groups (Landis, 1995; Hickerson, 2001, 11-12). Furthermore, large-scale and extensive data integration projects<sup>1</sup>, and the rapid increase in and availability of digital meta-data and object-data have diminished the extent of human influence and increased the necessity to automate steps, or at least to appraise the feasibility of such endeavours, which until now have been subject mostly to successful human deliberation. Internal challenges result from increased efficiency requirements; either because of budget cuts, increased user demands and expectations towards access and services (Craig, 1998; Frusciano, 2008b), or workloads due to growing masses of archival materials.

Standards for encoding archival descriptions for digital information systems, such as the de-facto standard for the digital encoding of archival aids, the *Encoded Archival Description*<sup>2</sup> (EAD), essentially translate the traditional structures and paradigms of archival aids to the digital realm and are thereby prone to perpetuating their shortcomings (Eidson, 2002). Evaluations of information systems implementing EAD encoded archival aids have raised doubts about the usefulness of such systems (Yakel, 2004) and pointed to what appears to be a *semantic gap* (Shaw, 2013, 1102) between what users want and what archives offer.

The digital age has also been considered a chance to rethink the concept of finding aids (Meissner, 1997; Gilliland-Swetland, 2001) and to develop new access tools and archival information systems that could even augment to “part finding aid, part expert system, and part intelligent agent able to conceptualize, mediate, and tailor the information provided” (Anderson, 2004, 113-114).

However, the main problem neither rests with the indirect approach to archival information access embodied in archival aids nor with issues of technical implementation but in the question concerning what knowledge is relevant for archive users in archival aids and digital archival information systems. The pivotal prerequisite for the improvement of existing digital archival information systems and the design of their future successors (Anderson, 2004) is overcoming the prevailing lack of qualitative in-depth analysis of archival user needs (Craig, 2003; Sinn, 2010).

Archives need to gain empirical confidence regarding the kinds of entities that should be documented and the relevant relationships that connect these entities. Ontological models of the information needs of archive users may provide the necessary theoretical and logical framework. Such an ontology needs to be firmly rooted in the needs of archive users and their conceptualizations and not be oriented only towards conceptual ideas from the archival side in order to be able to bridge the perceived semantic gap between user and archive (Doerr and Iorizzo, 2008; Shaw, 2013).

Without an adequate model of archival information needs the archival curator has also no objective guideline for assessing the relevance of the accumulated knowledge to the user. The

---

<sup>1</sup> For example, the archival Linked Data projects APENet (<http://www.apenet.eu/>) and APEx (<http://www.apex-project.eu/>), which are also closely tied to the Europeana (<http://www.europeana.eu/>) context.

<sup>2</sup> <http://www.loc.gov/ead/>

archival curator can only rely on intuition based on professional experience in order to decide which facts have sufficient relevance for particular kinds of research interests. This intuition mostly rests on human communication in terms of human understanding of textual entries in archival aids and therefore is highly implicit and informal. A guideline for assessing relevance would need to contain explicit and formal knowledge based on empirical evidence pertaining to the information needs of archive users.

Recognizing these problems, the archival profession has expressed the need to address the lack of user studies and to foster research in the domain of *archival information behaviour* as well as to re-intensify dialogue with their users, in particular the historical profession (Cook, 2011). Consequently, archive user studies have increased in number over the last few decades, with an increased focus on the users. While these user studies focus on many different aspects of the information behaviour of users in the context of archives, no ontological model of archival information needs has been created until now.

Furthermore, most of these archive user studies have been dominated by methods of data collection which lead to comparably biased analyses of archival information needs. Most user studies employ techniques which either primarily elicit intuition, such as interviews or questionnaires with users, or a particular reaction to an existing information system, such as observation or log file analysis (Doerr and Iorizzo, 2008). Consequently, the information needs to be drawn from user studies built on such interrogative or observational means are strongly biased by either a human mediator or a particular system setting.

Duff and Johnson (2001) conducted one of the few archive user studies which specifically collected and analyzed reference questions sent by email to North-American archives. *Archival reference questions* are written or oral natural language inquiries posed to archives which commonly constitute requests for information provision regarding the holdings of an archive or other kinds of service requests related to the archive's business.

This kind of empirical data allows for the establishment of a stronger and more genuine user perspective built on relatively unfiltered information needs. This is because users have formulated their questions in their own words, relatively independently from mediating parties (Duff and Johnson, 2001, 44), and prior to having received an answer to their questions. Therefore, such archive reference questions constitute valuable and promising empirical research data for the investigation into archival information needs.

### **Research Proposal**

The *interest* of this study is to remedy the lack of empirically based and ontologically formalized knowledge derived from qualitative in-depth user studies about archival information needs pertaining to state archives.

In particular, the study aims to (1) provide empirical insight into the nature of written user inquiries made of archives and their relation to archival documentation, (2) propose and apply a principal framework for the interpretative analysis of written natural language inquiries to

archives, and (3) devise a model to serve as a formal and explicit ontological representation of the subject matter that typically appears as the interest of reference questions to state archives.

The main *research question* inquires as to whether an ontology can be created that allows for a reasonable representation of archival inquiries and their probable interpretations as formal queries to the model of the archival target world that would adequately address the purpose of the inquiry.

The study entails three principal *hypotheses*. Firstly, it assumes that an interpretative analysis of archival reference questions in written natural language can serve to identify common patterns of shared interests towards archives and that the subject matter of these interests can be adequately formalized in an ontological model. Secondly, the subject matter of these interests exhibits dominant structures that do not disintegrate into a limitless richness of detail but can be highly generalized and abstracted to general ontological patterns. Thirdly, the ontology *CIDOC CRM*<sup>3</sup> (CRM) as one of the most successful in describing the cultural and historical domain to which the reference questions can be expected to pertain, provides adequate means for the ontological formalization of the subject matter and interest of reference questions users pose to archives.

As mentioned previously, the real questions of archive users are not those that are recorded in interviews or captured as queries to archival information systems. Nor are they the archival reference questions as such. The *real questions* of users posed to archives are to be found through interpretative and ontological analysis of written natural language reference questions. They materialize as the subject matter which typically appears as the interest of archival reference questions.

For this purpose, 762 written natural language reference questions were collected from two European state archives, the German Federal Archives (*Bundesarchiv*) and the National Archives of Norway (*Arkivverket*). A methodological approach was developed and applied, focusing on the interpretation and ontological modelling of the common and shared subject matter of users' interests. Formally, this study practices *knowledge engineering* in the domain of *archival information behaviour*, and employs as its method an *ontological analysis* of empirical written natural language questions posed to archives.

The methodological approach by Duff and Johnson (2001) has been the initial inspiration for the development of the interpretative approach for the study at hand. They investigated the types of reference questions users pose to archives as well as the *given* information as provided in the inquiries and known to the user, and the so-called *wanted* information required by the user and qualified by the given information. This approach, however, is yet further characterized by the idea of a keyword search and access, and does not allow for an ontological representation of archival information needs.

The interpretative approach developed for the analysis of reference questions extends the approach by Duff and Johnson (2001) and formally consists of two major steps which are

---

<sup>3</sup> <http://cidoc-crm.org/>

conducted iteratively: The first step remains at the level of utterance of the reference questions and determines the type of questions as well as the wanted information and given information. The second step then proceeds to the deeper interpretation of the interest of inquiries and common ontological entities constituting their shared subject matter.

The interpretation is primarily based on the most recognizable sense of the inquiries but also guided by an epistemological framework comprised of basic and essential constituents of the historical and archival domain as well as common knowledge about the world and background knowledge of the interpreter. This epistemological framework underpins the interpretative analysis of the inquiries and allows to make educated assumptions – where necessary – about the interest of the inquiries and potentially relevant responses for the user.

The principal *outcome* of the study and its most significant *contribution* to the current state of research on archival information behaviour is the *Archival Knowledge Model (AKM)* and the *methodological approach* to the analysis of archival reference questions. The AKM is an *ontological model* that evaluates, uses and extends the ontology CIDOC CRM as a means to formalize *general patterns*, which represent the subject matter typically appearing as the interest of inquiries towards archives.

These general patterns constitute the AKM and together formally and explicitly describe the information needs of archive users. Information needs of archive users thus formally and explicitly represented in the form of an ontology provide the necessary theoretical framework and conceptual foundation for closing the semantic gap between conceptualizations of users brought to an archive and the conceptualizations of archives offered to users through archival aids.

Furthermore, the AKM allows the analysis of existing archival information systems and knowledge bases as well as encoding schemas such as EAD, and inquires whether they provide sufficient semantics for serving typical archive user needs. Archivists will be able to assess and decide on the relevance of particular information based on empirical evidence provided by the general patterns of the AKM. New *cataloguing rules* may thus be derived, stating which information entities should be made explicit and how they should be related to each other. Finally, the model may lead to new *pattern-based query* (Dworman et al., 2000) mechanisms for archival information systems.

This study neither investigates the information-seeking behaviour of archive users in general, nor of historians and archivists in particular. Their search strategies and methodologies, the use and usefulness of obtained archival materials, questions of interface or system design, and implementation all exceed the scope of this study. No attempt is made to determine specific user groups. Rather, this study deliberately refers to the archive users as those users who have submitted reference questions to the archive.

Furthermore, the study is a qualitative study of archival information needs, and does not seek to provide statistics on the distribution of particular user needs. Similarly, while the interpretation and modelling represents a certain degree of plausibility, it is not logically binding.

This means that the AKM consists of plausible relationships and constitutes a meta-model or theory of the abstracted and formalized subject matter appearing as the interest of archival user inquiries.

The interpretative analysis interprets information in free text in terms of a formal data structure utilizing constructs from the ontology CIDOC CRM. The goal is not to describe or assume an algorithm that automatically converts free text either from user inquiries or existing archival descriptions – into a particular formal data structure. Rather, the study demonstrates that archivists are in a position to encode equivalents of the information found in archival reference questions in such a formal data structure based on the AKM. Still, progress in natural language translation suggests that at least semi-automatic conversion algorithms could be developed that immediately support archivists in the application and evaluation of the AKM and that render information systems informed by the AKM accessible.

Most importantly, this study does not aim to evaluate or critique archival principles and practices. Nor does the AKM seek to change or interfere with the ways in which archives are described and documented. The archival domain is ultimately the use case chosen for the study. Its results, namely the AKM, however, provide the theoretical framework for the evaluation and critique of archival description and documentation principles and with which to foster efforts towards better access facilities and data quality as well as data integration in the archival domain. The ultimate goal and purpose of the AKM is to improve recall by searching for plausible chains of relationships in data structures relevant to specific yet typical interests of archive users and thus to improve information retrieval.

## **Outline**

Chapter II provides a review of related research and will further contextualize the topic of the study. Chapter III introduces the empirical research data, which has been obtained from two state archives, the German Federal Archives and the National Archives of Norway. The characteristics of archival reference questions will be discussed, as well as the selection and collection process.

Since the development of a new method is one of the major objectives of the study, chapter IV introduces at greater length the methodological approach devised and applied to the analysis of the empirical research data. After revisiting the approach of Duff and Johnson (2001), the extended interpretative and ontological analysis of the study will be developed step-by-step. This will include a discussion of the historical and archival domain, both of which contribute to the epistemological framework, influencing the interpretation of the archival reference questions. The ontology CIDOC CRM as the prospective primary means for the ontological representation of the results of the interpretative process will then be introduced. The chapter concludes with a summarizing discussion of the methodological approach developed in the context of knowledge engineering.

Chapter V presents the results attained from applying the interpretative analysis to the



archival reference questions. The chapter reports on the types of questions as well as the identified given and wanted entities, and continues with a discussion of various results obtained during the interpretative analysis. These results both contribute to the empirical understanding of user needs and pertain to the development of the general patterns. The main part of the chapter focuses on a detailed presentation of the ascertained ontological representations of the primary subject matter, which typically appears as the interest of user inquiries posed to archives. These general patterns constitute the AKM and the primary outcome of the study and its contribution to a better empirical and ontological understanding of archival information needs. The chapter closes with a summary of the AKM.

Chapter VI investigates on an exemplary basis the relation of the AKM to real-life archival data as found in archival finding aids encoded in EAD. The conclusion (VII) summarizes the primary outcomes and contributions of the study and outlines future research. The appendix contains scope notes for all classes and properties that form the AKM.

## Chapter II

# Related Research

This chapter discusses related research that has been instrumental to the study at hand. Most of this research originates from the archival domain. In a necessarily simplifying discussion, the research of this study will be characterized as an investigation of archival information behaviour and located within information science.

The following three sections will then discuss studies on the information behaviour of archival users, beginning with research on archival information-seeking behaviour, followed by research on archival information needs and, finally, studies on the design and implementation of digital archival finding aids. The discussion will characterize the predominant topics and methods employed, and the self-conception of the archival profession regarding the state of research on archive users. The research topic of this study has been chosen in response to the identified shortcomings of the research on archive user needs. The chapter concludes with a discussion of the *Global Knowledge Network* (Doerr and Iorizzo, 2008), the research agenda of which aligns with the focus of investigation in this study.

## 1 Information Behaviour

The term *information behaviour* (Case, 2008), according to Bates (2010), denotes “the many ways in which human beings interact with information, in particular, the ways in which people seek and utilize information”. As a field of study in library and information science, information behaviour refers to “a wide range of types of research conducted in order to understand the human relationship to information” (Bates, 2010). More specifically, as Wilson (2000, 49) has put it, information behaviour comprises all human behaviour “in relation to sources and channels of information, including both active and passive information seeking, and information use”. The definition clearly points to two fundamental sub-fields of information behaviour: *seeking for information* and *use of information*.

While the study of use of information as well as of the ‘usefulness’ of information is important, the study at hand focuses on investigating seeking for information. The study of *information-seeking behaviour* constitutes a central part of the field of information behaviour. Wilson (2000, 49) has defined information-seeking behaviour as “the purposive seeking for information as a consequence of a need to satisfy some goal”. The complex process of seeking for information – often also termed as information searching – has been further characterized in different models; in the influential models by Belkin (1980), Kuhlthau (1991), or Ellis (1993; 2009), for example. While these and other models of information-seeking have been revised or further developed to advocate different perspectives, they nevertheless all share the fact that information-seeking follows information needs.

Information needs are to be understood as secondary order needs, which are the result of the desire to satisfy a primary need (Wilson, 2000, 51), and which signify “a consciously identified gap in the knowledge available to an actor” (Ingwersen and Järvelin, 2010, 20). According to Taylor (1962, 392), the term *information need* is further characterized by four levels of “the actual, but unexpressed, need for information (the visceral need); the conscious within-brain description of the need (the conscious need); the formal statement of the question (the formalized need); the question as presented to the information system (the compromised need)”. Such information needs then may initiate or be the basis of information-seeking activities (Ingwersen and Järvelin, 2010, 20).

The archival reference questions analyzed in the study at hand pertain to the third sense; that is, the formal expression of the information need by the user in the form of a written natural language question. The empirically grounded investigation of archival information needs constitutes the formal research focus. Studies investigating aspects of information behaviour may generally be called “user studies” (Wilson, 2006). As such, this study is an archive user study with a particular focus on the information behaviour of such users.

Harris (2005) has pointed out that it was not until the mid-1980s that the importance of archive user studies was emphatically emphasized, for example in articles by Freeman (1984), Conway (1986) or Dowler (1988). Conway (1986) suspected a methodological issue behind the reluctance of the archival profession to investigate user needs, and proposed a framework for the systematic and structured observation and analysis of archive users over a longer period of time. Such an approach would also allow for a systematic comparison of user studies based on the same methodological framework. Regular “reference logs” were designated for data collection on three principle topics: the user in general, the use of archival materials, and users’ information needs.

Later, Lack (2007) proposed four methodologies – focus groups, interviews, questionnaires and usability testing – for collecting input and feedback from archival users. Similarly, Proffitt (2007) has discussed various user-centred approaches for the continuous evaluation of user needs including a range of methods such as usability studies, focus groups, interviews, ethnographic studies, and web log analysis. In 2008, Anderson proposed a model for the systematic planning and evaluation of websites of online archives and their functionalities. Duff et al. (2008) has investigated how standardized questionnaires for user-based evaluations could be developed and which kinds of methods archives currently employ in order to gather user feedback on archival services.

Yet, until today, many archive user studies have ignored the necessity of additional and more diverse as well as more systematic and qualitative archive studies, which indicates a persistent deficit in that area. In 2008, Gray again deliberated the future role of archives as depending on further archival outreach, and additional research on building adequate services which would meet the needs and demands of growing and public user groups. Recently, Sinn (2010) reiterated that research on information behaviour in the archival domain would exhibit a lack of

qualitative in-depth user studies. Similarly, Trace and Dillon (2012) and further Hu (2012) have also advocated additional qualitative user studies and the application of new and more diverse methodological approaches to the study of archive users.

The following non-comprehensive discussion of archive user studies in the field of information behaviour will provide an overview of those studies that have been consulted.<sup>4</sup> Furthermore, the goal of the discussion is to provide an impression of the kinds of topics investigated and methods employed in archival research, as well as to delineate the critique ushered by the archival profession regarding their own approaches over the last two decades. The first section will present a range of archive user studies which have primarily focused on different aspects of information-seeking behaviour. It will then turn to those user studies that have specifically investigated archival reference questions and similar expressions of information needs. The last group of archive studies discussed will indicate that the apparent shortcomings in the available knowledge about archive users have also become evident in the design and implementation of archive information systems, in particular digital and online finding aids. These user studies mostly pertain to the information-seeking as well as information needs of archive users.

## 2 Archival Information Behaviour

Early studies of information-seeking behaviour in archives include Beattie (1989/90) who surveyed Canadian historians studying women's history. The study found that participants utilized informal means such as archivists, footnotes, and other colleagues alongside formal descriptive access tools such as finding aids. Beattie (1989/90, 47) further stressed the potential of subject-oriented finding aids and the need for more elaborated descriptive means of access to archival holdings. The study by Hutchinson (1997) also represents a relatively early report on an experiment investigating the recall and precision of retrieval operations in finding aids encoded in EAD. Participants executed four different search strategies in 250 online finding aids at the University of California, San Diego, and the University of California, Berkeley. They used 20 prototypical reference questions which had been considered relevant in the context of the available finding aids. Among several methodological and topical recommendations for future studies, Hutchinson (1997) points out the importance of employing dedicated user-based perspectives in user studies as well as the biasing influence that results from investigating users operating in pre-existing access systems.

Duff and Johnson (2002), stressing the importance of more qualitative user studies in the archival domain, studied the information-seeking behaviour of ten historians by conducting semi-structured interviews regarding their search behaviour for primary sources within an archive, how they conducted their research, and their use of archival materials. The study primarily focused on the means of the information-seeking behaviour, and identified four

---

<sup>4</sup> Most titles discussed in the current chapter have been collected from archival journals and by citation chaining.

involved activities: orientation, seeking known materials, building contextual knowledge, and identifying relevant material (Duff and Johnson, 2002, 492). The collection of names and generally the orientation towards institutions and persons who might have created relevant documents were important. The historical background knowledge of the researcher was further instrumental in this context (Duff and Johnson, 2002, 493-494).

Tibbo and Marchionini (2002) summarized the initial results of an international survey among 300 historians of American History in the U.S. on their information-seeking behaviour and how they locate primary sources in the digital age. The study found that most participants preferred traditional, non-digital means for searching though not denying newer digital ones such as electronic library catalogues or archival finding aids. Tibbo (2003) then reported on the final results, concluding that multiple pathways into the digital material are necessary and that users were wary regarding electronic search methods.

The information-seeking behaviour of genealogists constitutes the particular focus of the study by Duff and Johnson (2003) who conducted in-depth interviews with ten participants. The study found that genealogists preferred informal sources of information such as colleagues. Regarding information needs, the importance of person and place names was stressed, as well as the necessity to adopt archival finding aids to the needs of users.

The results of a postal questionnaire among 600 faculty members of history departments in Canada about their information-seeking behaviour in archives as well as the potential of digital technologies are the subject of Duff et al. (2004a; 2004b). Both studies conclude that digital technologies such as digital finding aids, digitization and email would facilitate and improve search and retrieval but that the encounter with the original document would not become superfluous. In the same year, Hamburger (2004) asked historical researchers about their search strategies when looking for primary sources. Traditional and analogue approaches were most important, along with searches based on person names.

The information-seeking behaviour in archives of UK academic historians is the focus of a study by Anderson (2004) who conducted a questionnaire in order to investigate the usefulness of digital finding aids and similar retrieval tools. The study is a good example of how the introduction of new digital technologies have not only incited more in-depth studies of user needs but, at the same time, unveiled some of the inadequacy of the existing paper-based finding aids. Anderson (2004, 83) points out the widely (and falsely) adopted assumption that the print paradigm of archival finding aids would translate into the digital age. According to Anderson (2004, 114), to “simply retro-convert print finding aids or create new ones in their likeness and place them all in an electronic environment” would not be an adequate strategy. Instead, considerations should lean towards new “on-line archival systems that are part finding aid, part expert system, and part intelligent agent able to conceptualize, mediate, and tailor the information provided” (Anderson, 2004, 114).

With the advent of the digital age and the realization of the increasing importance of online and remote users (Hill, 2004), archive user studies also began to specifically investigate the

digital and online information behaviour of archive users as well as the usability of digital and online services and resources of archives and online finding aids in particular. These studies have repeatedly stressed the need for additional and preparatory user studies during the design of digital archival services and information systems. Altman and Nemmers (2001) describes the development and implementation process of an online finding aid and search engine in the POLARIS project. The project conducted a dedicated case study which directly involved users and evaluated the usability of the online finding aid. The study ascertained a high level of interest of users in online finding aids and emphasized the value of user feedback for the development of archival information systems.

Rosenbusch (2001) compared twelve websites providing access to archival finding aids and found that user needs were not sufficiently considered. The prospective user groups as well as adequate contextual information remained unknown due to an “almost exclusively supply-side driven” (Rosenbusch, 2001, 44) approach to the design of online archival finding aids. Consequently, Rosenbusch (2001) deemed user studies on information needs and information-seeking behaviour a prerequisite for creating effective online archival services.

The study on online finding aids conducted by Hostetter (2004) utilized interviews and questionnaires and found that online finding aids were considered useful in principle, while their particular advantages and usefulness remained unclear. One of the conclusions of the study was that stronger involvement of users as well as participation of archivists in standardization efforts of archival metadata standards were needed.

Cruikshank et al. (2005) has analyzed and discussed the situation of description and digitization of archival materials and whether current practices in archival description utilize the potential of digital and online technology to address the needs of users. While issues of presentation and website design, archival jargon, navigation, accessibility of search and browsing functionality, as well as providing better instructions and guidelines to users had already been identified by previous studies, the question as to which kinds of information would be useful to users remained unclear. Similarly, Peters (2005), who reported on the GASHE project suggested adding full-text information on activities and functions to the description of archival records in finding aids. The traditional interpretation of archival principles such as provenance and original order alone would not be sufficient to appropriately cover the whole context of archival holdings.

Cox (2008) investigates how the utility of printed and digital finding aids for professional users could be analyzed with a non-archival perspective in order to overcome the archive-centric epistemological viewpoint in most archival studies. The usability study by Chapman (2010) analyses user interaction with online finding aids and also indicates deficiencies of current archival information systems such as that the “aboutness” of archival materials should be described more thoroughly and archival jargon hinders users in their understanding of archival descriptions.

Nimer and Daines (2008) conducted user studies to gain a better understanding of the user

perspectives and information needs towards archival collections and how digital technology, here the Web 2.0, could help to improve digital finding aids. The study has been conducted as a usability study as part of an archival project whose main objective is to improve the “usefulness” of an archival information system by introducing a concept model as an additional layer on top of the finding aid and tailoring it towards typical user needs and tasks. The study by Nimer and Daines (2008) is a good example of utilizing digital technology, here a conceptual model, as a supplement and not a replacement to existing archival aids.

The rather new form of digitally-born and electronic archival records has been broached as an issue by, for example, Grimard (2005) who stresses the importance of their long-term preservation, or Adams (2007) who looks at the use and users of data records at the National Archives and Records Administration (NARA) in the U.S.<sup>5</sup> Bailey (2007) has urged the archival profession to distinctively include digital records and contents in archival theory. However, this study does not specifically consider these new forms of archival records.

### 3 Archival Information Needs

Dowler (1988) has emphasized the importance of investigating not only the actual use of archival holdings but also the potential use by focusing on the users and, most importantly, on their questions as well as the nature of inquiry itself. There are several user studies which focus on archival information needs by analyzing (reference) questions, queries to archival information systems, and similar expressions of information needs. Such studies are particularly important for the modelling of archival data and designing appropriate metadata schemas as well as in determining relevant information to be presented and made available to archive users.

Only few studies have specifically looked at reference questions posed to archives. The most important was conducted by Duff and Johnson (2001) who investigated 375 email reference questions collected from various North-American archives. The focus of their analysis rested on the type of questions asked in these reference questions as well as their structure; that is, the information desired – also called the “wanted information” – and the information given in these inquiries. Duff and Johnson (2001, 44) underlines the importance of analyzing natural language questions to archives and that by investigating “what elements people use to describe their information need and how they structure their requests will make it possible to design better research aids, including automated information retrieval systems, finding aids, and web interfaces”. The method employed by Duff and Johnson (2001) adapts a methodology for analyzing library reference questions based on the work by Grogan (1992) and Jahoda and Braunagel (1980) and will be further discussed here in the “Methodological Approach” chapter (IV). The results of the study by Duff and Johnson (2001) are mostly in accordance with previous, comparable studies. They ascertain that user inquiries mainly focus on resource

---

<sup>5</sup> <http://www.archives.gov/>

discovery (material-finding, specific form and known item), service requests, user education, administrative/directional, fact-finding and consultation, while the most frequent terms used in questions include proper names, dates, places, subjects, forms and partly events. General subject terms were less important.

The study of Maher (1986) is an early example of a cursory analysis of 159 written reference letters to the archive of the University of Illinois. "Object of inquiry" questions about specific subjects and specific documents dominated over inquiries for "photograph or film", "specific record series or collection", "specific data", "general subject", "general holdings information" and "genealogical information". In sum, more than 40% of all reference letters appeared to be "relatively narrow requests with clearly stated goals" (Maher, 1986, 20).

Bearman (1989/90) states that one of the main obstacles on the path to designing better information systems was the blurry knowledge about the "kinds of questions" asked and the actual "content of questions". The study conducted by Bearman (1989/90) collected written and oral questions of users via a questionnaire in several archives during one day. Then the presentation language was analyzed in terms of the type of questions and the access points provided. The most important elements in user questions were form, personal and corporate names, citations and titles, places, topical subjects, and dates (of subjects) (Bearman, 1989/90, 4-5). In contrast to oral questions, written inquiries had the most access points, according to Bearman (1989/90), and would often provide indications of the research interest (Bearman, 1989/90, 6).

In a large-scale study, Conway (1994) analyzes how researchers communicate with archives and the elements of research questions. The most frequently used elements were date, place, medium, personal name, and broad subject. Conway (1994) identifies four major categories of archive users: academic, occupational, avocational, and personal. Each of them accommodated several user groups which exhibited different information-seeking and information-use practices derived from the information needs and methods of their particular disciplines.

Gagnon-Arguin (1998) collected questions posed via telephone, email or in person to archives in Quebec. The study pointed out the value of user queries as an information source about the user. The pivotal information elements for search strategies identified in this study were "general subject", "place", "date", "event", "occupation" and "form".

Collins (1998) investigates the user needs related to retrieval in image archives. A survey collects questions posed to two photograph archives. The most frequently used elements were general and specific subject terms, followed by dates and place names. The study stresses the importance of subject indexing and subject access for digital image archives. Users did not ask about the context of images, which is probably due to the fact that almost no inquiry was part of a historical research project. Similar work from related domains include, for example, Chen (2001) who investigated image inquiries from 20 college students, or Choi and Rasmussen (2003) who studied user queries regarding image retrieval in American history. Further, Fear (2009) looks at the terminology used in Dublin Core metadata fields in historical image collections.



The study by Martin (2001) collects a large sample of 595 correspondence units by letter, telephone, facsimile and e-mail sent to the Southern Historical Collection and General and Literary Manuscripts at the University of North Carolina at Chapel Hill. While the analysis focuses primarily on reference questions as communication channels to archives, however, it does not specifically investigate the structure of those questions.

Related research has been conducted in the context of subject access, which has been discussed as an additional means of access to archival records. For example, an early study regarding subject indexing, which included a retrieval experiment, has been conducted by Lytle (1980a,b). Pugh (1982) discusses the application of subject access and retrieval in an archival context. The study noted, however, that far more insight is needed into the archive users, their information needs and the use of archival resources. Smiraglia (1990) describes how the Library of Congress Subject Headings would be used for subject access to archival materials and proposes a methodology for subject analysis and cataloguing. Dooley (1992) argues that traditional ideas of access by provenance to archival material are no longer sufficient, and discusses the prerequisites and the necessity of subject access and subject retrieval for archival information systems. The adequate design of subject access could only be determined, however, by conducting additional user studies investigating who uses the archives and their level of satisfaction. More recently, Fenton (2010) has discussed the use of controlled vocabularies and subject keywords for indexing archival finding aids.

The investigation of reference questions and queries to archives regarding information needs has not been in the focus of archival research until now. No study has analyzed the semantic relations between the elements within reference questions. The results obtained are helpful but still pertain to the idea of keyword searching. The fragmentary nature and general persistent deficit in knowledge about archive users and their information needs is indicated by user studies focusing on EAD and respective archival information systems.

## 4 Archival Implementations of EAD

Deficits in the systematic and in-depth analysis of user needs in the context of archives are also evident in the design and implementation of the *Encoded Archival Description* (EAD) standard, which has received special interest in the domain of archival information behaviour. EAD has become the international de-facto standard for the encoding of digital finding aids (VI:1). The success of EAD, of course, has been a big step forward towards improving international standardization and cooperation (Kiesling, 2006) as well as digital access to archives.

However, Eidson (2002) has pointed out that user needs were missing in the initial design of EAD as well as other archival descriptive standards. Similarly, Carter and Frusciano (2004) wrote in the editorial of the *Journal of Archival Organization* that user needs had been neglected during the development of EAD. There was still a great demand for more in-depth user studies, especially studies focusing on online finding aids and digital sources. Coats (2004) conducted a

literature review on user studies of archival finding aids in general and finding aids encoded with EAD in particular. The report pointed out that additional and broader user studies of finding aids encoded with EAD were necessary as a prerequisite for appropriately assessing their adequacy as well as for further developing the standard.

Early studies on EAD include Meissner (1997) who reports on the analysis and evaluation of traditional (paper-based) finding aids by the Minnesota Historical Society in order to offer online service and access to their archival descriptions. The report stresses the importance of first assessing and possibly revising finding aids before moving on to the application of EAD. Re-engineering of finding aids should constitute the first step before moving to online representations and services, on account of many issues such as unclear identification of information elements, unsatisfactory arrangements of information elements, alternating descriptive layers, and missing user guides.

Other examples of early studies specifically involving EAD are Duff and Stoyanova (1998) who examines the content and display of online finding aids in EAD, or Szary (2001) who discusses the potential of EAD particularly for self-sufficient user communities and more productive archivists. Roth (2001) explores the deployment and evaluation methods of EAD finding aids and suggests that users and their feedback needs to be more involved in the process. Similarly, Eidson (2002) stresses that user needs should be considered more strongly during the further development of EAD.

EAD patterns have been analyzed by Prom (2002b) and the “EAD Cookbook” as a guideline for their encoding by Prom (2002a). The results indicate that many EAD finding aids are inconsistent and lacking in key elements. Both studies recommend better best-practice guidelines for their encoding with a view to optimizing their usability and retrievability. Redding (2002) also examines the adaptation and implementation of EAD finding aids, and discovers the need to uphold the quality of their contents and to further standardize their encoding and annotation in order to meet the modern terms of data-centric modelling.

Kim (2004) has studied a variety of archival finding aids encoded in EAD regarding their usability and recommends better navigational aids, browsing and navigational functions as well as controlled access points via person names, place names, or general topics. The usability study of EAD interfaces by Yakel (2004) indicates that users generally demand better interfaces with less archival jargon and that do not demand prior understanding of archival hierarchical documentation traditions. Yakel further identifies issues with search functions and content display.

Zhou (2007) has evaluated search functions of archival websites with EAD finding aids and found that search access is still limited, also due to the lack of user studies in this area. Similarly, Huffman (2008) has conducted an analysis of the features and functionality of EAD retrieval systems using a content-analysis approach, concluding that EAD does not fulfil its potential. An XML retrieval approach has been presented by Zhang et al. (2009), intended to demonstrate how EAD finding aids could be exploited for better search and retrieval. Zhang and Kamps

(2010) have also carried out a transaction log analysis of archival information systems analyzing the queries submitted by specific user groups to EAD archival finding aids. They identified a significant difference between the search behaviour of expert and non-expert users.

Dow (2009) conducted a review of the state of EAD and concluded that it qualifies as a “halfway technology”, due to its complexity high cost factor, and the high level of adaption it requires from archivists. Halfway technologies “address symptoms of a problem but not the causes or long-term effects” (Dow, 2009, 108). The crucial issue is to better connect researchers with the archival materials. The key question here is whether EAD can help to solve this problem (Dow, 2009, 109-110). EAD has been and continues to be adapted, yet “ultimately, however, much more sophisticated solutions lie in the future, perhaps involving applications that fall under the label of Web 2.0, Web 3.0, and applications we have not yet imagined” (Dow, 2009, 110). Dow (2009) concludes that “the EAD elements that we encode today, properly and fully done, will become the raw materials of any data-centric high technology that replaces it” and that if “properly done now, our EAD encoded finding aids will not need redoing later” (Dow, 2009, 114).

The study by Gilliland-Swetland (2001) once again stresses that the potential of EAD does not rest with the replication of the physical and intellectual form of finding aids for online distribution but in the chance to re-conceptualize finding aids and to prepare them for new application scenarios, especially in terms of discovery functionalities. Gilliland-Swetland (2001) adapts the “berrypicking” approach by Bates (2009) and outlines a couple of search capabilities to enhance browsing and retrieval in EAD-based archival finding aids.

The general idea of re-conceptualizing, extending, or supplementing finding aids with additional knowledge layers or conceptual models in order to enhance access and to reconcile conceptualizations of different user groups has been entertained not only by Gilliland-Swetland (2001) or Nimer and Daines (2008) as discussed above, but also by Yakel (2004), who discusses the idea of archival access tools as potential boundary objects. Such boundary objects “are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites” (Star, 1989, 46). EAD, understood as such a boundary object and not only as a mere data structure, “must not only mediate between archivists and their user communities, but must also facilitate a convergence between the user and the archival content” (Yakel, 2004, 64) and thus even act as boundary spanners during research processes. Broad implementation, acceptance, and employment by the archival profession and user would be a prerequisite for reaching that goal, according to Yakel (2004). The same idea has been discussed by Cox (2008).

Recently, Sinn (2010) has examined how archive users perform research on a specific topic in an archival collection and how they perceive the importance of the found material for their work. The results show research patterns which differ from the typical assumptions of archivists, thus indicating a divergence of cognitive representations. This suggests that there is in fact no need to change EAD itself, but rather a boundary object which would close this gap.

The *Archival Knowledge Model* (AKM) could help to achieve this goal by providing an ontological representation of archival information needs against which EAD and archival aids encoded in EAD could be evaluated. While this study does not investigate or analyze EAD itself, nor does it examine its adequacy regarding user needs, it develops the means to do so. The studies presented here illustrate the current state of research in the archival domain with regard to EAD and user needs.

## 5 Summary: Global Digital Access

To summarize the discussion on related archival research, the general tendency appears to be that current implementations of archival information systems are not the best method with which to render archival descriptions and archival materials accessible to users, as indeed Anderson (2004) has stated. Furthermore, as Cox (2008) has pointed out, implementations and services utilizing EAD seem not to have significantly improved archival access. The implementation of archival information systems based on EAD is, of course, also often confronted with technological barriers as well as the lack of readily available expertise and resources (Yaco, 2008). However, one of the primary issues remains the uncertainty when it comes to users and their information needs.

Furthermore, most user studies have employed either observational or interrogative means in order to collect research data. Data collection through techniques such as the observation of user interactions or the analysis of log files mostly only elicit reactions to particular and existing information systems and databases (Doerr and Iorizzo, 2008). The information need of the user is strongly filtered in terms of what the user is either allowed or able to do within the boundaries of the particular system or what the user thinks the database may be able to answer (Lytle, 1980a, 69).

Similarly, when addressing users directly, for instance by interview or questionnaire, and asking about the questions they would want to ask of a particular information system or database, the users will provide questions they think the database might be able to answer based on a particular mental model (Gentner and Stevens, 1983). The same is true if asking users outside of a specific working process which kinds of questions they might have with regard to a specific system. The users will mostly rely on their intuition because for psychological reasons users cannot properly remember the original questions they had in mind before receiving a reply, the original question having since become obsolete.

Less attention has been paid to the investigation of information needs and to explicit user perspectives during user studies. In particular, requirements of metadata and data structures that would meet archive user needs in terms of search and retrieval operations appear to have been neglected. User studies specifically analyzing reference questions and similar expressions of information needs also seem rare, despite the fact that reference questions do promise relatively

unfiltered information needs.<sup>6</sup>

Not long ago, Schaffner (2009) conducted a literature review on the needs and requirements of users with regard to archival information systems and other similar infrastructure. The study concluded that sufficient knowledge on the users existed and that this knowledge only needed to be applied. Schaffner (2009) saw the key to meeting user needs in metadata design and creation. With the previous discussion of archival research in mind, the former statement seems to be not quite correct, at least regarding the archival domain. The latter statement, however, that the enriching of metadata with contextual historical knowledge and insights will add further meaning to (archival) objects (Boonstra et al., 2004, 103) and thus better cater for user needs, does stand to reason. Yet in order to design such appropriate metadata and respective archival information systems, further empirical insights into the information needs of users are necessary in order to determine which contextual information is relevant to them.

As Cook (1984, 46) has put it, “archivists must transcend mere information, and mere information management, if they wish to search for, and lead others to seek, ‘knowledge’ and meaning among the records in their care”. Similarly, Case (1991, 79-80) has expressed doubts as to whether scholars, and especially historians, have been served well by disciplinary boundaries imposed by traditional (library) classification and indexing schemes. Even though changing documentation principles or reordering collections has evidently not been feasible, the so-called *problem-oriented model* (Weintraub, 1980, 29) would provide a more relevant framework for devising better tools and services by shifting the focus to “points of view” or “contexts” for indexing. Today, one of the key challenges of organizing information is “to construct systems that aid understanding, contextualizing, and orienting oneself within a mass of resources” (Shaw, 2013, 1100), and where models help to bridge the *semantic gap* between the formalizations in information systems and the conceptualizations of scholars.

An ontology – a concept which will be introduced in detail in chapter IV – may provide this kind of additional “point of view” and “context” in the form of a description of those general and common things which are important to users when consulting an archive. Such an ontology is a means of communication and point-of-reference by explicating the meaning of these things and, most importantly, how they relate to each other. The context created by relationships is pivotal for an understanding of culture-historical objects since “understanding is built on associations” (Doerr et al., 2007, 51). The archive user studies discussed above exclusively conceptualize user needs and generalize information needs in terms of single entities. At the same time, these studies neglect to elaborate on the relations between these entities and their further contexts. This study aims to illustrate and describe these interrelationships.

Users, archivists, and system architects alike may use such an ontology as a reference when consulting, evaluating or planning archival descriptions, metadata schemas, or archival information systems. Another advantage of ontologies lies in their independence from particular

---

<sup>6</sup> Diaries and abstracts might also be sources of such relatively unfiltered information needs. The latter have been investigated by, for example, Tibbo (1993).

implementations of data structures such as the graph-based *Resource Description Framework*<sup>7</sup> (RDF) or the tree-based *Extensible Markup Language*<sup>8</sup> (XML). For this reason, this study refrains from directly applying the results of the analysis of reference questions to metadata standards such as the XML-based EAD.

Adequate ontologies may also eventually help to address one of the core problems of today's information systems on the Web, which is the limitation to "simple answers to clear cut, search term-based questions" (Doerr and Iorizzo, 2008, 1). Pattern-oriented retrieval (Dworman et al., 2000), for example, by exploiting relationships between entities, could describe far more complex questions whose answers move beyond the capacity of simple queries. This limitation poses a significant barrier to more sophisticated and integrated information systems.

The Semantic Web (Antoniou et al., 2012) in particular promises to address these problems; however, its research agenda suffers from an "almost exclusive focus on 'terminology' rather than 'ontological structures'" (Doerr and Iorizzo, 2008, 4), a trend which could also be attributed to the body of archival research discussed above. The prevalent focus on traditional library cataloguing and methodology in describing and contextualizing objects of interest also contributes to these limitations. As a consequence, important complementary lines of research have been neglected such as the investigation of the nature and structure of original user questions. The systematic and in-depth analysis of original user questions from different stages of the research process, however, is essential and has the potential to provide, for example, necessary information on query mechanisms and on an adequate granularity of ontologies (Doerr and Iorizzo, 2008).

The basic and central question, therefore, is which contexts are relevant to users and how these contexts can be reasonably represented in an ontological model. The scrutiny of related research as detailed above leads us to the proposal of contextual descriptive knowledge for archival finding aids that is derived from empirically based qualitative user studies of information needs. Instead of employing observational or interrogative means for data collection, reference questions are collected and analyzed which promise to provide relatively unfiltered information needs in the users' own words. The analysis should assume a user perspective and describe the user needs in an ontological form emphasizing the various interrelationships.

First, however, the next chapter will discuss the archival reference questions that were selected and collected as research data. Then the chapter "Methodological Approach" (IV) will introduce the interpretative analysis and further discuss the concept of ontology.

---

<sup>7</sup> <http://www.w3.org/RDF/>

<sup>8</sup> <http://www.w3.org/XML/>

## Chapter III

# Empirical Data

Archival reference questions constitute the primary source of empirical data for this study. As already stated in the chapter “Related Research” (II), such questions contain largely unfiltered information needs, which renders them particularly interesting subjects for empirical analysis.

In this chapter, the nature of reference questions in the context of the archive and their particular value as empirical research data will be further investigated. Then, the data collection and selection process for the two samples obtained from the German Federal Archives<sup>9</sup> (*Bundesarchiv*), hereafter also “BArch”, and the National Archives of Norway<sup>10</sup> (*Arkivverket*), hereafter also “NAN”, will be explained. Single inquiries extracted from these two samples constitute the actual unit of analysis. This first step of the empirical data analysis will be discussed in the third section. The summary of this chapter will reflect upon the inherent bias of any such sample of archival reference questions.

## 1 Reference Questions

Archives offer *reference services* (Duff and Fox, 2006) in order to provide information about or from their holdings, thus rendering the latter available to researchers (Cox, 1992). The term *reference question* denotes a request made by a user to a staff member of a library or archive for information or assistance regarding the provision of some kind of information. Such requests can be either placed in person, for example at a service desk in a reading room, or remotely by telephone, post or email.

Reference questions may come from administrative bodies, research institutions, individual researchers or private persons. There is a broad range of purposes they may serve, such as identifying the first reference to a particular historical place in medieval documents, genealogical questions, recent personnel issues, property disputes, or any kind of historical interests including legal, economic, ecclesiastical, or art-related matters (Franz, 2004, 111-112). To a certain degree, an archive often serves as a kind of historical information office, even in the absence of departments officially designated as such and despite the fact that the institutional rules of procedure may not specifically allow these kinds of services.

Even though questions posed in the reading room or via telephone can also be deemed reference questions, normally no primary documentation of such conversations exists. Additionally, oral inquiries evolve during a reference interview in which the staff member attempts to narrow down and identify the information need of the user by conversing with them. During this process, the original information need of the user changes and becomes biased by the questions

---

<sup>9</sup> <http://www.bundesarchiv.de/>

<sup>10</sup> <https://www.arkivverket.no/en>

of the interviewer. Even if primary records of such conversations existed, these records would have to delineate in detail how the initial inquiry of the user shifts during the reference interview. Furthermore, as shown in the literature review, applying observational or interrogative methods during data collection also leads to research data containing biased information needs of users.

Written reference questions, on the other hand, are primary and empirical research data which represent a largely unfiltered information need expressed in the user's own words. The users depict and record their conceptualizations of their research interests relatively independently of the direct intervention or immediate influence of a third party. The additional contextual and background information provided by the users in their inquiries is deemed relevant by them and constitutes personal background knowledge that is known to them.

Written reference questions are, of course, not without influencing factors and do not exhibit a completely unbiased and original information need. For example, users may have had a telephone conversation prior to having formulated a letter or e-mail inquiry, or the user may have sent an email from the reading room just after speaking with an archivist. In general, the user relies on personal background knowledge and experience while composing an inquiry. This background knowledge and experience evolves or changes during the user's work within the archive and the further exploration of the research question.

As part of the service, written reference questions are collected in so-called *user files* in which archives consolidate the correspondence and other business transactions with their users. These user files typically contain, in chronological order, printouts of emails or copies of letters sent to the archive by users, copies of the replies from the archive, various administrative forms regarding the service provision of the archive to the user, and background material from research performed by an archivist in response to an inquiry.

As such, user files constitute a highly information-rich source of primary research data for analyzing user needs from various different research perspectives. Depending on the extent of the records, user files would, for example, allow investigation into the evolution of information needs during the user's engagement with the archive or shed light on internal procedures and communications during the processing of user inquiries by archivists.

The main advantage for the study at hand, however, is seen in the fact that these written inquiries present the information need in the user's own words since there is no immediate factor directly influencing the user while formulating an inquiry. The reference questions can therefore be expected to contain an information request that is to a large degree unfiltered and further qualified with contextual information known to and considered relevant by the user.

The user files are the point of reference during the initial collection and selection process of the two samples. Physical and digital copies of user files were collected and selected by information systems providing access to user file records. These information systems can be used in order to select and retrieve sets of user files based on a range of search facets depending on the particular system, such as titles of user files, dates of reception, or the purpose or topic of the inquiry.



In the following two sections, the collection and selection processes for the two samples of user files will be discussed in greater detail.

## 2 Data Collection

The first and main sample of user files originates from the German Federal Archives (*Bundesarchiv*) and the second from the National Archives of Norway (*Arkivverket*). As the Federal Archives of Germany, the *Bundesarchiv* was chosen primarily because large quantities of genuinely historical inquiries were to be reasonably expected, and in broad variety. Furthermore, the German Federal Archives showed support for this project from an early stage, which facilitated the process of obtaining access to its reference services. The geographical situation of the German Federal Archives was also convenient to the author, allowing for work on-site with the original copies of user inquiries and permitting easy consultation with reference archivists. The National Archives of Norway were chosen because they are also a state archive and comparable to the German Federal Archives in terms of size and significance. The National Archives of Norway also provide easy online access to digital copies of user inquiries.

The only pre-determined condition for the collection and selection process for the first and main sample from the German Federal Archives was that the user files and the questions they contained were to focus on a similar historical and topical horizon. Contemporary history, understood here as essentially European history of the 19th and 20th centuries, was selected since this field in particular is confronted with a rapidly-growing abundance of mostly written remnants of the past which constitute the empirical basis of any historical inquiry (IV:2.1) into recent history. Due to their sheer quantities, these pose a considerable challenge to the historian seeking relevant and adequate source materials. A further reason for the selection of this historical period was the author's own familiarity with this specific epoch, which would be of advantage in the analysis of the inquiries (IV:2).

The sample of user files to be collected from the German Federal Archives as well as the National Archives of Norway was to contain as many substantial questions as possible. The term "substantial question" loosely refers to any kind of inquiry that would generally be relevant and adequately phrased to be processed by an archival information system or the expertise of an archivist, such as questions aiming at obtaining archival materials or factual, historical information. Questions about administrative matters such as opening times, or service inquiries such as requests for copies are not the target of the analysis. The particular categories of questions will be introduced and discussed in the next chapter, "Methodological Approach" (IV:1.1.1). This study does not focus on the statistical distribution of different types of questions; rather, it conducts an in-depth analysis of substantial inquiries such as research-related, material-seeking or fact-finding questions and aims to extract shared interests.

## 2.1 The German Federal Archives

As a *state archive*, the German Federal Archives are responsible for the permanent preservation and accessibility of federal documents such as files, papers, cartographic records, pictures, posters, films, sound recordings and machine-readable data. Historically, the German Federal Archives have to date maintained the civil and military archives of the Federal Republic of Germany (since 1949), and of its predecessors, the German Confederation (1815-1866), the German Reich (1867/71-1945), the occupation zones (1945-1949), and the German Democratic Republic (GDR) (1949-1990). It also holds documents from important institutions of the Holy Roman Empire (1495-1806). In addition, important documents of private origin, and from political parties, associations and societies are kept in the German Federal Archives. In particular, as a special department of the Federal Archives, the “Foundation Archives of Parties and Mass Organisations of the GDR” (*Stiftung Archiv der Parteien und Massenorganisationen der DDR im Bundesarchiv*, SAPMO) maintains the archives of the central leadership organs of the parties, unions and mass organizations of the GDR.

The German Federal Archives granted supervised access to their *user files* at its Berlin branch in Berlin-Lichterfelde. The user files contain a physical documentation of the correspondence between a user and the archive. Typically, this documentation comprises email printouts or copies of user inquiries and archivists’ replies filed in chronological order and containing various internal remarks, user management forms and miscellaneous administrative matter and, occasionally, material from research performed in response to an inquiry. The German Federal Archives receive some 60,000 written inquiries per year.<sup>11</sup> In principle, for every user who contacts the German Federal Archives a user file is created, in which all subsequent correspondence with this user is collected.

Each user file has an identification number, retrievable through two different internal databases. MachMIM is used in the register of the German Federal Archives to record every analogue or electronic inquiry submitted to the archive. BASYS-2 is used in the reading room to organize information related to users who visit the archive in person. Such information includes, for example, which archival materials were ordered by the user to the reading room during his/her visit. Both databases allow the researcher to search for identification numbers of user files.

BASYS-2 was used to retrieve a sample set of user files because it offers a small range of additional search facets related to the users and their associated user files. This includes, for example, the general purpose of the inquiry as given by the user on the user management form. The user may choose from a given set of general purposes such as research for a dissertation, press, film, television, genealogy, publication, or local history. Other search facets include a general subject and time-frame of the inquiry’s topic, or the department initially responsible for processing the inquiry. Since the various departments of the German Federal Archives focus on

---

<sup>11</sup> These numbers are based on the public report of the German Federal Archives from 2008 and 2009: <http://www.bundesarchiv.de/oeffentlichkeitsarbeit/publikationen/taetigkeitsberichte/>

particular historical provenances, knowledge of the department allows the researcher to infer the origins of the archival material and the interest of the initial inquiry of a user.

However, it is important to note that these classifications are coarse and not specific enough for a precise retrieval of user files based on these search facets. For example, users may indicate several purposes of inquiry on the user management form or none at all. The initially responsible department for processing an inquiry may change and other departments may participate in the process or take over responsibility. The correspondence between the archive and the user may also contain or relate to different topics over time. None such changes are coherently recorded in the database. Nevertheless, and in contrast to MachMIM, BASYS-2 enables limited but basic control over the selection of user files and for the creation of a qualified and purposeful sample. The sample of user files was therefore selected purposefully by employing the following search facets in BASYS-2.

During the first search iteration, the facet “responsible department” was limited to the value “StA”.<sup>12</sup> Limiting the searches to this facet permitted a focus on inquiries primarily related to contemporary German history, more specifically related to the history of the German Democratic Republic (GDR), and thus a topic-related focus regarding the inquiries in the user files.

The facet “purpose of the inquiry” was used to place a further limit on the result lists to ensure, on the one hand, rich user files and, on the other, a degree of variance. The values included “dissertation”, “genealogy”, “exhibition”, “publication”, “press”, “illustrated book”, “edition”, “film”, “television”, “local history”, “qualification thesis”, and “diploma”. Between 10 and 20 identification numbers of user files were selected from each result list, depending on the assumed richness and value of the user files.

A second search iteration was performed with the value “R1”<sup>13</sup> for the facet “responsible department” in order to further supplement the sample with potential inquiries falling into the same, yet earlier part of the shared topical and temporal horizon. Limiting the two search iterations to “StA” and “R1” for the facet “responsible department” allowed for the collection of a relatively broad sample of user files. Furthermore, another important aspect was the physical availability of user files of these two departments at the Lichterfelde branch of the German Federal Archives. On the one hand, the selected user files could be expected to fall into a shared general topical and temporal horizon; on the other, they necessarily represent only a selection of all possible questions posed to the German Federal Archives. The second sample collected from the National Archives of Norway was also meant to partly address this issue and to alleviate the natural bias in the sample.

The selected identification numbers were then given to employees of the German Federal Archives, who checked whether the chosen user files complied with the imposed access conditions. Access to user files was mainly restricted by the following factors: (1) No user files of governmental bodies were allowed; (2) users must have agreed to the proliferation of their topics

---

<sup>12</sup> Department “StA” primarily holds files of political parties and mass organizations in the GDR.

<sup>13</sup> Department “R” holds files from Imperial Germany (*Deutsches Reich*).

on a user management form; (3) each user file was carefully perused by an archivist in order to ensure that none contained sensitive or personal information. User files cleared for access were then made available for this study in reading room.

Thus, 196 user files were selected by performing several searches utilizing different search facets in BASYS-2 and based on educated assumptions after scanning the selected user files. The sample was further complemented by a special selection of 40 rich user files. These had been collected by the head of the department SAPMO, and pertained to the chosen temporal and topical horizon. They were therefore highly likely to contain substantial questions.

Altogether, 236 user files were selected. Access was granted to only 100 user files, however. This was because either the file did not comply with the previously stated access restrictions, or because the user file was currently in use by an archivist. Of these 100 available user files, 60 contained at least one question or request. One of the main reasons for the high number of user files not containing any questions or requests is due to the fact that users frequently establish contact by phone or, in fact, visit the reading room without making previous contact at all. In such cases, a user file is created which often contains only a user management form and other administrative matter. User files do not contain transcripts of telephone conversations.

The user files collected from the German Federal Archives share a general historical and topical horizon that focuses on German contemporary history, understood as the history of the 19th and 20th centuries. The BASYS-2 information system allowed the selection of user files likely to contain relevant questions. In the case of the first and main sample from the German Federal Archives this was a conscious choice, based on the assumption that similar patterns of research interest are more probable in questions relating to similar historical and topical horizons.

There was no reliable data available regarding the types of users and their backgrounds, although the users' professions were provided. The value and thus reliability, however, of such information is questionable as, for example, the user may profess to be an academic historian while in fact working in a journalistic capacity. For that matter, as it is certainly possible for a lay historian to reach a level of proficiency exceeding that of an academic historian, the question of a user's profession is of marginal importance.

In sum, the sample from the German Federal Archives comprised 60 user files that contained overall 227 single emails or letters. On average, each user file contained 3.8 single pieces of correspondence.

## **2.2 The National Archives of Norway**

In order to validate the results from the first and main sample, a second sample was collected from the National Archives of Norway. The aim was to collect a purposeful set of 60 case files each containing at least one substantial question from a random base set. The second sample of case files was collected after the initial analysis of the first and main sample of user files from the German Federal Archives. The aim was to collect a comparable set of empirical research data,

both in terms of quantity and quality, in order to validate the findings from the first analysis.

The National Archives of Norway are comprised of the National Archive (*Riksarkivet*) in Oslo and the regional state archives. Like that of the German Federal Archives, the primary task of the National Archives of Norway as a state archive is the permanent preservation of archival material from national and regional Norwegian state institutions and to enable and ensure public access. As such the National Archives of Norway specifically presents itself as “administration bodies and scientifically based institutions for culture preservation”. The preserved records include “the non-current records of government ministries and other central offices”, the oldest dating from 1189 A.D. The National Archives of Norway also preserve records from private companies, organizations, political parties and private persons.<sup>14</sup>

The archive receives about 50,000 incoming inquiries annually.<sup>15</sup> Most inquiries are sent to the National Archive (*Riksarkivet*) in Oslo directly. The correspondence with users is collected and documented in *case files* and recorded in an internal journal system. The case files correspond to the user files in the German Federal Archives and contain copies both of the inquiries sent by users and of the replies from the archives along with material collected during research on user inquiries. After reception and registration, inquiries are forwarded to the relevant departments and processed by an archivist assigned to the case. As in the German Federal Archives, indirect or broad questions are covered depending on the resources available.<sup>16</sup>

The records on case files from the internal journal system are accessible via the *Electronic Public Records*<sup>17</sup> (*Offentlig Elektronisk Postjournal*, OEP) database, which went online in 2010. The OEP is a searchable general database that makes available public records on inquiries sent to the various governmental and administrative bodies in Norway. Digital copies of inquiries can be requested and are sent via email, which greatly facilitates access to case files. Since the OEP system thus greatly simplifies the identification of potentially relevant case files, and because of the convenient accessibility to digital copies of single inquiries, the National Archives of Norway were chosen as the source for a second sample of reference questions from a non-German national state archive. Furthermore, since the National Archives of Norway can be seen as the Norwegian counterpart to the German Federal Archives, similar and comparable types of inquiries were to be expected.

The OEP system describes cases and single documents. Cases are comprised of one or more single documents pertaining to one topic or area of interest of a user. A document is either an email or a letter. Each case and document has a title that consists of semi-controlled terms assigned by the attending department.

The advanced search allows for several search restrictions such as keywords to appear or not

---

<sup>14</sup> <https://www.arkivverket.no/en/about-us/the-national-archives-of-norway>

<sup>15</sup> The seemingly high annual number of inquiries as compared to the 60 000 inquiries of the German Federal Archives is probably due to the fact that the National Archives of Norway comprise nearly all state archives in Norway, including regional ones. In Germany, many more state archives exist as independent institutions.

<sup>16</sup> The information provided here is based on an email exchange with the National Archives of Norway.

<sup>17</sup> <https://www.oep.no/>

appear in the case titles or document titles, various date restrictions related to the reception of the inquiry, entry into the internal journal system, and the publication of the inquiry in the OEP database, the name of the institution which received the original inquiry, or whether the document was incoming (sent by user to archive), outgoing (sent by archive to user), or internal. However, the search options do not offer any semi-controlled thematic, topical or temporal search facets related to the content of the inquiry as in the case of the German Federal Archives. The contents of a case or single document can only be approximated by interpreting their titles.

In addition, most case and document titles, as well as the contents of inquiries for that matter, were of course in Norwegian. *GoogleTranslate*<sup>18</sup> was used to overcome the language barrier. Preliminary tests showed that the automatic translations provided by GoogleTranslate were sufficient to understand case and document titles and, more importantly, to recognize the sense and research interest of the contents of inquiries. In cases where words could not be translated or the meaning was not obvious, native speakers were consulted.

Furthermore, the OEP database currently<sup>19</sup> contains nearly 240.000 records relating to the National Archives of Norway alone. At the same time, the system can display only 10 hits per result page. Limiting the search result by setting it to incoming inquiries only; that is, excluding all messages sent out to the user in response to an incoming inquiry and all internal communications within the archives, still amounts to nearly 130,000 hits. Since it is not possible to download or otherwise automatically process the records in the system, and manually clicking through the result set on a 10 hits per page basis would have been inefficient, the result set was further reduced to a smaller set of potentially relevant cases and documents.

For this purpose, the study employed the following advanced search criteria. The agency from which the record originates was set to “National Archives of Norway”. The document type was limited to only *incoming* messages, and the recipient of these messages was set to *Riksarkivet*. The two last filters prioritized inquiries sent to the main branch of the National Archives of Norway in Oslo, which is mostly equivalent to the German Federal Archives since the records of the Norwegian governments and of most national administrative bodies are kept at this location. Accordingly, inquiries comparable to those collected from the German Federal Archives were to be expected.

The filter applied provided a set of keywords<sup>20</sup> to be excluded from the document titles. This was possible since the document and case titles are assigned by the attending authority in a quasi-controlled manner. These keywords were collected iteratively during the first assessments of the OEP system and in consultation with employees of the National Archives of Norway. Even though the extent and consistency of the keyword assignments remains unclear, the exclusion of

---

<sup>18</sup> <http://translate.google.de/>

<sup>19</sup> As of September 11th 2014.

<sup>20</sup> The keywords included the following terms: “*Eiendom*” (property), “*Kopi*” (copy), “*Bestiller*” (order), “*Jobbsøknad*” (job application), “*Innsynsbegjæring*” (access request), “*Innsynsbegjæring*” (disclosure petition), “*Søknad*” (application), “*Praksis plass*” (internship), “*Personalmappe*” (personal folder), “*Dødsfall*” (death), “*Skifte*” (change), “*Adopsjon*” (adoption), “*Jobbmulighet*” (job opportunity), “*Skanning*” (scanning), “*Kopibestilling*” (orders for photocopies), “*Kassasjonsbestemmelser*” (cassation provision), “*Deponering*” (disposal).

these specific keywords from document titles helped to eliminate vast numbers of inquiries that would not be of interest for this second sample, due to the focus on user files including at least one substantial question.

The last filter applied was the date on which the user sent the inquiry to the *Riksarkivet*. The selection process was then performed in two iterations. During the first, the date filter was set to the time-span “from 01.01.2012 to 31.12.2013”, which resulted in 7276 total hits. The first 2500 hits were explored by reading the document and case titles on every 5th result page, where each result page contained 10 hits. The *complete* case file was ordered if (1) any case title or document title on a given result page suggested that a case file or single document would contain a substantial question, and (2) free access to the case or document was not restricted by privacy regulations. The complete case file would generally provide complementary information to the user files of the German Federal Archives. In this way, 55 case files were ordered during the first iteration.

In order to maximize the coverage and randomness of the sample, a second iteration was performed during which the date filter was set to the time-span “from 11.09.2013 to 11.09.2014”.<sup>21</sup> The first 750 hits of the 7916 total hits were explored in the same way as during the first iteration. The search was stopped earlier, however, because a sufficient number of case files (i.e. 55) had already been identified and ordered.

The sample was supplemented by random but purposeful picks of five additional case files from the initial exploration of the functionality of the OEP system, and seven more case files forwarded by a Norwegian researcher consulting on the process of identifying those which were relevant. These seven case files were included since they contained substantial questions and therefore were adequate for the purposes of the second sample. They were re-checked for completeness.

All together, 123 case files were ordered and received from the National Archives of Norway. This sample was analyzed in the order of receipt of the digital copies. Only 76 case files were analyzed before 60 case files were found to contain substantial questions. The other 16 case files analyzed did not contain substantial questions or requests.

Other than in the case of the first and main sample from the German Federal Archives, the sample from the National Archives of Norway was collected without any intention to exclusively focus on questions concerning contemporary history. This was in order to create an open-ended second sample and to retain the potential to compare the first sample with questions possibly relating to very different historical and topical horizons. The question remained whether the impartial interpretation of questions from a different, albeit similar, archive would result in the same patterns as did the interpretation of the first sample. Topics related to other historical eras were therefore not excluded while assessing titles of case and document titles. However, the final second sample, even with its slightly broader topical variance, mostly contained questions related to contemporary history.

---

<sup>21</sup> Corresponds to the option *last year*.

Information about the users is not available in the OEP database. The contents of the inquiries only provide a vague, if any, indication on the background of the user. However, as with the German Federal Archives' sample, most users appear to have either a personal historical interest in a certain topic, or they make inquiries within an academic context.

In sum, the sample from the National Archives of Norway comprised 60 case files which contained overall 96 single emails or letters. On average, each user files contained 1.6 single emails or letters.

### 3 Single Inquiries

Up to now, the term inquiry has been only loosely employed in order to generally refer to emails, letters, or any kind of oral request, all of which may contain any number of reference questions. From this point on, the term *inquiry* is more specifically defined as a single direct or indirect (reference) question that is typically further qualified by associated contextual information. As already mentioned, only *written* inquiries posed either by letter or email were analyzed in the study. Letters and emails may contain any number of single inquiries. The principle units of reference during the analysis of the empirical data are single inquiries. After the collection of the two samples with user and case files from the German National Archives and National Archives of Norway, therefore, the necessary preparatory step before the analysis was the extraction of single inquiries from the files.

The following example presents a single inquiry from a case file obtained from the National Archives of Norway. The case file consisted of an email, Figure 1, and an attached letter, Figure 2, sent to the archive.<sup>22</sup> In this particular case, the inquirer is forwarding a letter previously submitted to another authority, which then referred the user to the National Archive of Norway.

[REDACTED] - [REDACTED]

Fra: [REDACTED].co.uk]  
Dato: [REDACTED] 21:53:58  
Til: riksarkivet arkiverket  
Tittel: [REDACTED] - [REDACTED]

---

Dear National Archives

I sent the attached letter to the Sjøfartsdirektoratet in Haugesund, but unfortunately they were unable to help and suggested I contact you. I hope you will be able to help.

With thanks in anticipation of your reply.

Yours very sincerely

[REDACTED]

Figure 1 – Email sent to the National Archives of Norway.

<sup>22</sup> Names of persons, address information, and other personal data have been rendered unreadable.



This example shows that the collected inquiries may have been already submitted to institutions other than the German Federal Archives or the National Archives of Norway. This example also demonstrates that inquiries and the associated contextual information may appear in separate locations.

Tel: [REDACTED] - [REDACTED].co.uk

[REDACTED] 2013

Dear Maritime Authority

**Norwegian [REDACTED] - [REDACTED]**

We hope you can help with our search for information about the award of this medal.

[REDACTED] was [REDACTED] grandfather. His medal is now in our care and will eventually be passed to our sons. [REDACTED] sailed on fishing vessels from [REDACTED] became first-mate, and died at sea, on [REDACTED] as a result of enemy action during the [REDACTED]. We understand from family stories that, sometime in his fishing career, he was involved in helping to rescue some Norwegian sailors/fishermen.

We have discovered that his medal must have been awarded before [REDACTED], since it bears the portrait of [REDACTED], who is described as [REDACTED]. On the reverse is the inscription [REDACTED]. The medal is silver.

We would be really pleased to learn if any further details exist, at this distance in time, about the events which led to the granting of this award.

Yours very sincerely

[REDACTED]

Figure 2 – Letter attached to an email containing an inquiry with one question and associated contextual information.

The two pages contain one single inquiry consisting of one question with associated contextual information. In this particular example, the actual inquiry is provided in the attached letter. The question is stated twice in indirect form at the beginning and end of the attached letter. The user asks for any information about the events that led to the granting of an award to his wife's great grandfather. The associated contextual information is provided in the remainder of the letter, including the subject matter line printed in bold right after the salutatory address.

All emails and letters in the user files and case files in both samples were read. Parts of the text that, for example, introduced the author of the inquiry or gave further personal background were eliminated. Inquiries were extracted and recorded in a table if an indirect or direct single

question or request was identified. The associated contextual information relevant to a single question, given anywhere in the email or letter, was added to the question in the table. During the extraction of single inquiries, no differentiation was made between discrete and follow-up questions. Every identifiable single implicit or explicit question was treated as a single question in its own right even if the questions were similar. In the rare cases that questions were essentially identical, or only differed slightly in terms of phrasing as shown in the previous example, the repeated question was not treated individually.

The delineated preparatory process of extracting single inquiries from user and case files was in itself already an act of interpretation. Determining whether two similar questions are indeed identical or whether they exhibit sufficient distinct meaning may serve as an example for interpretative challenges. However, in most cases, common sense sufficed to decide which parts of an email or a letter constituted a single question and which contextual information was relevant. The extraction process was further facilitated since most questions in a letter or email refer to the same topic. The result of the extraction process was recorded in a table structuring all single questions and associated contextual information provided by a user in an email or letter.

Of the 60 user files from the German Federal Archives, 564 single inquiries were extracted as described above. The 60 case files from the National Archives of Norway resulted in 198 single inquiries. In total, the sample for this study comprised 120 user and case files, 60 from each archive, from which 762 single written inquiries were extracted. This sample of single inquiries constitutes sufficient empirical research data for sound statements related to the research questions of this study. To the author's best knowledge, no comparable sample of written archival reference questions in terms of quantity and quality exists or has been analyzed to date.

## **4 Summary: Sample Data**

Archival written natural language questions posed by mail or email conserve a relatively unfiltered and genuine information need relating to an archive, compared to research data collected via other observational or interrogative means. The two samples of user and case files containing such reference questions were collected from the German Federal Archives and the National Archives of Norway. The data collection process, as previously described, ensured a large degree of randomness while still retaining a shared historical and topical horizon in both samples relating to contemporary history (European history of the 19th and 20th centuries, mostly of Germany and Norway). Additionally, since the point of reference during the data collection was the user and case files, both containing random sets of inquiries, no potential type of question was consciously excluded. Analyzing and comparing two samples provided additional evidence for the validity of the results and helped to ease the natural bias inherent in both samples (Ioannidis et al., 2014).

The identification numbers of user files in the German Federal Archives include person names and therefore already carry sensitive personal information. In order to allow reproducibility and reanalysis (Ioannidis et al., 2014) of the results of this study, therefore, lists of the identification numbers of the collected user files have been deposited in the registry of the German Federal Archives and in the OEP system of the National Archives of Norway respectively.

Even though a substantial number of 762 single inquiries were extracted from the user and case files, both samples necessarily represent only a particular snapshot of all possible inquiries to state archives. However, since the inquiries mostly share an identifiable common scope, the results of the analysis can be expected, with near certainty, to be relevant and adequate at least to this particular domain of discourse covered by the inquiries.

Furthermore, the ontological model to be created based on the analysis of the inquiries contains entities that describe and represent what comprises the common and shared subject matter that appears as the interest of the inquiries. The more general and abstract these entities turn out to be; that is, the less specific they are to small subsets of the whole sample, the higher the probability that these structures are indeed relevant beyond the scope of the particular domain of discourse and empirical material.

Another important aspect of the sample refers to the dates on which the inquiries were sent to the archives; that is, the dates found on the copies of emails or letters. In the case of the German Federal Archives, the earliest inquiry was sent in 1999 and the latest in 2011. The inquiries in the sample from the National Archives of Norway were sent between 2009 and 2014. The inquiries in the whole sample, therefore, and the interests expressed in the questions, cover a time-span of about 15 years. The analysis of the inquiries will take into consideration that the interests of users may not only significantly differ geographically (that is, inquiries posed to German and Norwegian state-archives may have divergent interests); they also change diachronically.

Finally, the significance of this study does not solely rest on the specific results of the analysis of the inquiries, i.e. the ontological model, but is also found in the particular methodological approach employed here, which will be introduced in the following chapter.

## Chapter IV

# Methodological Approach

This chapter introduces the methodological approach devised for the purposes of this study and applied during the analysis of the research data presented above. Throughout this chapter, the methodological approach will be developed in a step-by-step portrayal.

The chapter “Related Research” (II) has shown that a semantic gap exists between the information needs of users and the information provided by archival information systems, and that approaches employed for the analysis of archival user needs exhibit particular biases. The aim of this study is to alleviate this semantic gap by augmenting the knowledge available about the information needs of archival users, and by contributing to the design of better digital archival information systems. The hypothesis here is that written natural language inquiries provide largely unfiltered information needs which can be interpretatively analyzed in order to identify the subject matter that typically appears as the interests of inquiries. This can then be formalized by means of an ontology.

An *ontology* – a particularly expressive kind of knowledge organization system to be further introduced throughout this chapter – can be understood, for now, as formalized knowledge comprised of clearly defined concepts and relationships concerning the possible states of affairs in a knowledge domain or domain of discourse (Stevens et al., 2010). By providing such clearly defined and agreed upon concepts and relationships as points of reference, ontologies allow a reconciliation of differing conceptualizations found in a general domain of discourse.

In the context of this study, this common ontological framework strives to provide a reconciliation between conceptualizations of archive users, as found in their inquiries, and those of the archival domain, as provided in archival aids and formalized in schemas and data structures as well as in the implicit background knowledge typically available to the archivists at the time. The knowledge obtained by the archivist during the documentation time is typically not exploited, and is either lost or only implicitly encoded in full-text fields. However, this knowledge could be utilized in order to create potential archival aids with a more effective structuring of the same knowledge without an increase in research effort. The challenge of identifying knowledge relevant to the user, however, remains. The investigation and formalization of the subject matter of the interests typically found in user inquiries made of archives, is the primary desideratum this study moves into focus.

One of the major advantages of an ontological formalization is that the manifold inquiries correspond to an ontology that formalizes their shared discourse and substantially generalizes and normalizes the diverse phenomena exhibited in the inquiries. As such, an ontology allows for the development or analysis of query schemas, and for the evaluation of their usefulness to information systems in relation to the original questions.

The primary research question here, then, was whether an ontology can represent archival

inquiries and their interpretations as formal queries to the model of the archival target world that would adequately respond to the interest of the inquiry.

The overall working hypothesis of the methodological approach distinguishes between two levels in the *interpretative analysis* (IV:1) of written natural language inquiries: the level of the utterance of the inquiry and the deeper interpretation of the intended meaning, its sense, and possible adequate answers to the question.

In the first section, these two principle levels of the interpretative analysis are developed by introducing and criticizing the foremost *linguistic analysis* by Duff and Johnson (2001) and then proceeding to an extended *ontological analysis* which was devised during and for the purposes of the study. While the first section delineates the overall approach of the interpretative analysis, the point of departure and its objective, the subsequent sections will go into more detail regarding the ontological aspects of the methodological approach.

The second section takes a step back and turns to the wider epistemological context of the interpretative analysis, presenting the *epistemological framework* (IV:2) pertaining to the historical and archival domain, both of which underpin and guide the interpretative analysis.

The third section discusses ontological aspects of the interpretative analysis by depicting the step from the immediate level of inquiry to the ontological representation of the subject matter of the interests of these inquiries as an act of *ontological modelling* (IV:3). Afterwards, the ontology chosen as the primary means to represent the results of the interpretative analysis, the *CIDOC CRM* (CRM), will be examined. Since the CRM is the most comprehensive ontology to describe the cultural and historical domain, albeit with a museum bias, the CRM is subjected to an unbiased examination of its general adequacy for the current domain in order not to reinvent concepts, in particular high-level ones. Finding it generally adequate, new concepts will be added as natural and compatible extensions as the interpretative analysis proceeds, providing evidence of such.

Formally, a *knowledge engineering* and *ontology engineering* (IV:4) approach is employed in that the common conceptualization identified in inquiries through interpretation are translated to an ontological model by evaluating and utilizing appropriate entities taken from the CRM or by creating compatible new ones. The possible extension of the CRM with compatible and adequate ontological entities is one of the explicit goals of this study.

## 1 Interpretative Analysis

In this section, the overall *interpretative analysis* of the inquiries will be developed and explained. Determining what users want to know when formulating a question and further providing contextual information as well as the underlying common ontological concepts is necessarily an act of interpretation. In order to stress this fact, the overall methodological approach has been termed interpretative.

Formally, the interpretative analysis conducted here can be broken down into an initial part

oriented more clearly towards the linguistic level of the inquiries and a subsequent part focusing on the underlying ontological background that forms the shared subject matter of the interests of the inquiries.

The initial linguistic approach is based on and inspired by the work of Duff and Johnson (2001) and will be discussed first. Afterwards, the ontological analysis will substantially extend the linguistic analysis and constitute the primary methodological focus of this study.

## 1.1 Linguistic Analysis

Linguistic methods for analyzing natural language sources have not yet been canonically categorized in the relevant literature. The purpose of a linguistic analysis of natural language sources is to identify and extract patterns that can be represented in formal models and implemented as algorithms for computation. During information retrieval, such qualitative linguistic methods promise better recall and precision than quantitative approaches such as statistical and probability-based methods.

In the context of information retrieval, according to Salton and MacGill (1987), there are three methodological steps identifiable in the linguistic analysis of natural language sources. The morphological method looks at words and analyzes their form, type and order. The syntactical method looks at sentences and analyzes their grammatical components. The semantic method considers the document as a whole and interprets the meaning and sense of the text in the light of contextual knowledge. In the context of this study, linguistic analysis is generally understood in the sense of work at the semantic level (Löbner, 2003) of the immediate meaning of words in a phrase, similar to the material addressed by automatic translation.

Duff and Johnson (2001) conducted a primarily linguistic analysis of archival reference questions, focusing on the syntactic level of the inquiries.<sup>23</sup> Their linguistic analysis is divided into three steps, which form a not strictly linear but iterative process. The *first step* constitutes the extraction of single questions and associated contextual information as inquiries. This step has already been performed and discussed in the previous chapter “Empirical Data” (III). The extraction resulted in 762 single written natural language inquiries obtained from 120 user and case files respectively.

The *second step* determines the type of the single questions, while the *third step* investigates so-called “Givens” and “Wanted’s”; that is, the information known to the inquirer and the information desired by the inquirer. These two steps are discussed in the following two subsections.

### 1.1.1 Types of Questions

After the extraction of single questions and associated contextual information from the reference questions, the *second step* of the linguistic analysis comprises the categorization of the single

---

<sup>23</sup> For the results of their study see “Related Research” (II:3).

questions according to different *types of question*. The primary purpose of this categorization is to separate questions inquiring after archival or non-archival materials or factual information from those containing other kinds of requests. Only the former two general types of questions – so-called resource-discovery and factual questions – are relevant to this study since only they pertain to the information found in archival finding aids.

Duff and Johnson (2001) devised a categorization schema for types of questions based on Grogan (1992), who specified a schema for types of reference questions asked in libraries.

They adapted and revised the schema by Grogan (1992) for the archival domain. The two main categories of question types are *limited help questions* and *open-ended questions*. In the schema by Duff and Johnson (2001, 49-51), the first category contains the question types *administrative / directional, known item, fact-finding, material-finding, specific form, and service request*. The second category includes the question types *consultation* and *user education*.

For the methodology used here, the schema by Duff and Johnson (2001) has been slightly extended and revised in scope and definition. The final types of questions utilized, and which will be introduced next, emerged from the iterative analysis of the inquiries. They fall into general categories: The first category, *resource discovery*, comprises questions that either explicitly or implicitly inquire after any kind of archival or non-archival<sup>24</sup> materials. The second category comprises *fact-finding questions*. The third category covers any kind of *non-discovery questions*.

The specific types of questions within each category will be introduced next.<sup>25</sup>

### Resource Discovery Questions

The group titled resource discovery questions includes *material-finding, specific type, and specific item* as questions that inquire *explicitly* after archival or non-archival materials, and *research question* and *consultation* as questions that inquire *implicitly* into archival or non-archival materials. In the case of implicit resource discovery questions, the user does not explicitly formulate a request for archival or non-archival material, but the inquiry implies a need for material. This interpretation is further justified by the fact that the German Federal Archives and, to a lesser extent due to resource constraints, the National Archives of Norway attempt to treat such implicit inquiries as resource discovery questions and to provide pointers to potentially relevant archival or non-archival materials.

---

<sup>24</sup> Examples of non-archival resources are secondary texts or documentary sources such as finding aids.

<sup>25</sup> All following examples have been translated from German to English by the author. Additionally, in accordance with German law, examples have been rendered anonymous by replacing personal names, place names, dates, topics etc. with categorical terms. Furthermore, the examples provided at this point omit any associated contextual information and only show the individual questions. Finally, each question has an identifier, assigned during the analysis. The identifier starts with the letter “Q” for “question” followed by three running numbers: the first number refers to the user/case file, the second one refers to the letter or email within the user/case file, and the last refers to questions within the letter or email. The subscripts “barch” and “nan” stand for German Federal Archives (*Bundesarchiv*) and the National Archives of Norway (*Arkivverket*) respectively.

**Material-Finding** Questions of the type *material-finding* directly or indirectly inquire as to the location or existence of *any* kind of archival or non-archival materials related to, about, or by, for example, a person or group, activity, or general topic or subject. In most cases, the user does not explicitly distinguish between primary and secondary or bibliographic resources but simply gives generic designations such as “anything about”, “something related to”, or “pointers to”. The user typically further describes the information need with related, contextual information about persons, places, or events, etc. This category is congruent with *material-finding* in Duff and Johnson (2001).

Q010-01-01<sub>barch</sub> “I am interested in archival material involving the former leader of the [political party name], [person name]. He led the party from [date]. Can you investigate whether there are more papers in your archive regarding [person name]?”

Q028-03-01<sub>nan</sub> “I wonder if there is anything short [sic] about the father (...)”

**Specific Type** Questions of the type *specific type* focus on a *type* of archival or non-archival material. The term “type” is used as a general and broad category subsuming various more specific terms including, for example, *forms* such as photo, microfilm, or copy, *functions* such as petition, survey, or personal letter, or *archival designations* such as record, personal estate, or (archival) collection. Specific type questions are less generic than material-finding questions since the requested material is designated by a specific type in the inquiry. This category is similar to *specific form* in Duff and Johnson (2001).

Q046-01-01<sub>barch</sub> “We are interested in any personal files. The person in question is [person name], née [person name] and born on the [date]. According to my information she went to school in [city name]. She also attended journalism school at the university [name of institution].”

Q044-01-01<sub>nan</sub> “Our hope and the question is whether there are drawings of [name of church]?”

**Specific Item** Questions of the type *specific item* give a name or designation of a particular single item or set of items, or are able to provide a detailed description. The item or set of items are known or assumed to exist. In this kind of question the user is even more precise about the requested item than in specific type questions. The user may also provide a specific type but clearly demand a particular item. This category corresponds to *known item* in Duff and Johnson (2001).

Q055-01-01<sub>barch</sub> “Could you please help me find the ‘Aryan certificate’ of my father?”

Q001-01-01<sub>nan</sub> “We (...) hope the National Archives can help us with a copy of the (...) War Children Committee settings that were in effect from [time span].”



**Research Question** Questions of the type *research question* state a primary research question which is also often specifically designated as such. Research questions convey an implicit need for relevant archival or non-archival materials, and are mainly characterized by the fact that “mere searching no longer suffices: what are [sic] needed are the more specialized tools of research” (Grogan, 1992, 42). Answers to such questions are essentially open-ended and mere searching for a fact or single documents does not suffice but, instead, a range of factual information or sets of documents is required in order to adequately answer the question. In particular, questions about the reasons and motivation behind certain acts, or how and why particular activities or events occurred, belong to this category. By implication, the user does not expect a direct and factual answer but typically pointers to potentially relevant archival or non-archival materials where a response or information may lead to further answers. This category equals *research enquiries* in Grogan (1992) and loosely corresponds to *consultation* in Duff and Johnson (2001).

Q005-04-13<sub>barch</sub> “*In what way did the FRG politically steer the preliminary investigations?*”

Q001-01-02<sub>barch</sub> “*One guiding question of my project is, for example, how the SED assessed the ramifications of the international process of deliberation towards the domestic politics of the GDR.*”

**Consultation** Questions of the type *consultation* reference a specific archival item such as a record or file, and directly or indirectly consult the archivist’s domain expertise or general historical knowledge about the archival item and its archival or historical context. Consultation questions often relate to the contents of particular records, or the inquire about the probability of finding certain types of materials or factual information within a specific archival holdings or files. In contrast to research questions, consultation questions clearly focus on particular archival items and express a clear indirect need for archival material. Furthermore, even though many consultation questions might appear to be existential questions, the assumption is that the user would not be content with a plain “yes” or “no” answer. Note that the type *consultation* in the current study is *not* equivalent to *consultation* in Duff and Johnson (2001) and only very loosely relates to their *user education*. For example, Duff and Johnson (2001) also subsume *research questions* under *consultation* while here they form a distinct type of question.

Q058-11-01<sub>barch</sub> “*I would like to know if those files contain the lists with the delegates of the congresses of the KPD.*”

Q001-02-01<sub>barch</sub> “*Do you know if materials of this working group have been archived with the files of the 'Ministerrat'?*”

These five types of questions characterize those questions that primarily aim at the discovery of any kind of archival and non-archival materials. The different types give rise to conclusions on various levels of specificity and scope of the inquiries.

## Fact-Finding Questions

The second main category of types of questions are questions of the type *fact-finding*, which correspond to the same type of question found in Duff and Johnson (2001). The user provides a specific and self-limiting question and expects the archivist to give a specific and factual answer. The decisive criterion for this kind of question is that the user does not ask for archival or non-archival material where an answer to the question might be found. The response is expected to be a factual answer containing specific items of information, including existential and truth confirmations.

Fact-finding questions do not exhibit a primary need for archival and non-archival material. Although consultation questions in some cases appear to be factual they always clearly imply the need for archival material. Similarly, research questions are not self-limiting by nature and also primarily imply a need for archival and non-archival material, even though possibly alongside factual information.

Q005-04-10<sub>barch</sub> “*Is it true that Adenauer sent his son-in-law to Bordeaux in order to defend the culprits?*”

Q026-01-06<sub>nan</sub> “*But also (...) the deployment locations of the police battalions would interest me.*”

Arguably, some fact-finding questions might turn out to be research question statements or will be treated as material-finding questions by the archivist. In such cases, for example, pointers to documents where an answer might be found are provided instead of a direct factual answer. However, if at the level of the utterance a question primarily asks for a specific set of facts and does not fit any other type of question then it is treated as fact-finding.

Finally, a wealth of questions describe unspecific information needs such as “anything about a particular person” or “something about a particular event”. Even though such questions might be satisfied by a range of factual information, since they are neither self-limiting nor requesting particular facts, and since the primary function of archives consists in providing access to archival material and not facts themselves, such questions are treated as material-finding questions.

## Non-Discovery Questions

The third main category of types of questions are *non-discovery* questions. These comprise questions that do not inquire after archival or non-archival material or request factual information. Non-discovery questions can be categorized into questions of the types *administrative or directional*, *service request*, and *user education*.

**Administrative or Directional** Questions of the type *administrative or directional* inquire about the archive’s administrative procedures or policies regarding, for example, opening and closing hours, access to and costs of photocopying services, the location of the reading room,

details regarding the accessibility and permitted usage of archival material, or the availability of specific service providers such as transcription services. This category also includes requests for appointments to use the reading room. This type corresponds to *administrative / directional* in Duff and Johnson (2001).

Q039-06-01<sub>barch</sub> “Am I allowed to display copies of documents from your archive in an exhibition?”

Q005-02-03<sub>barch</sub> “How long in advance do I have to book a place in the reading room?”

**Service Request** Questions of the type *service request* demand a typical service from the archive in its role as an archival institution; that is, to carry out a specific service. Examples include making photocopies of an archival document, ordering archival materials to the reading room in preparation for a visit, or mediating contact to a third party, for example, to the owner of a record for access permission or to a contemporary witness. This type corresponds to *service request* in Duff and Johnson (2001).

Q008-01-01<sub>barch</sub> “Please provide me with a letter confirming the feasibility of my research project with regard to the holdings of the Bundesarchiv.”

Q015-04-02<sub>barch</sub> “Could you please prepare parts of the materials in advance of my visit?”

**User Education** Questions of the type *user education* request education or guidance for the user in terms of technical, procedural, or methodological aspects regarding the usage of archives or research in general. Questions of the type *user education* typically contain only a very vague, if any, indication of what the user is looking for in the archive. In many cases, such questions are general statements of intent or ask how or where to get started with the research; in particular, how to use finding aids or how and where to perform searches. This category only *partly* corresponds to *user education* in Duff and Johnson (2001) who subsumes *research questions* under *user education*. Here, *research questions* form a distinct type of question.

Q049-01-01<sub>barch</sub> “I would like to know whom I can contact in order to begin with genealogical research.”

Q006-10-01<sub>barch</sub> “Is there a list or system in the finding aid which can be used to get an overview of the files which are on film?”

Again, even though non-discovery questions are coded during the initial determination of the types of questions, they are not subject to a further analysis of the Givens and Wantedes or the ontological analysis. Only resource-discovery and fact-finding questions constitute potential queries to archival finding aids and are therefore of interest.

The third step in the linguistic analysis of Duff and Johnson (2001) analyzes the information entities provided, the Givens and Wantedes, in the inquiries.<sup>26</sup> While the general idea of this

---

<sup>26</sup> Note that Duff and Johnson (2001) included all types of questions in their analysis of information entities.

third step has been adopted in this study, its particular conceptual design has been revised, as will be discussed next.

### 1.1.2 Information Entities

In the approach by Duff and Johnson (2001), the *third step* in the linguistic analysis is the identification and categorization of *information entities* in the individual questions and associated contextual information. The term *information entity* denotes the type of information represented by words or expressions used in an inquiry. For example, “Adenauer” is a personal name and “Bordeaux” is a place name. The approach used by Duff and Johnson (2001) stems from the work of Jahoda and Braunagel (1980), who proposed assigning every information entity in a question to one of three basic categories – Wanted, Givens, or Supplementary Information – and then to specify the type of information entity.

*Wanted*s denote the type of information entities that the user does not know and desires to obtain through the inquiry. For example, the user wishes to know the location of a document, the name of a specific institution or, more generally, find biographical information related to a specific person. At the same time, in the case of non-discovery inquiries, *Wanted*s may also refer to requests such as an appointment with an archivist or an order request of files to the reading room of the archive.

*Wanted*s remain a broad and unspecific category which can often only be determined by high-level interpretation when the goal of a question is not stated explicitly. Consequently, the most prominent *Wanted*s identified by Duff and Johnson (2001) are *biographical information*, *general information*, and *form*.

The *Givens* denote the type of information entities that the user provides in a single question and associated contextual information in order to describe and qualify the *Wanted*s. For example, a user may give the proper name of a specific person along with his or her date of birth and place of birth in order to find a birth certificate.

The schema devised by Jahoda and Braunagel (1980) also includes *supplementary information* as a third category, which designates information entities not vital for the identification of the *Wanted*. For this reason, Duff and Johnson (2001) do not include *supplementary information* in their data tabulation although they do encode this kind of information in their data analysis.

In this study, the methodological approach adopts the fundamental idea of identifying *Wanted*s and *Givens* in inquiries. However, in the context of the further ontological analysis of the inquiries, the character of the entities denoted by both categories is distinctively revised as ontological conceptualization. Here, this study shifts away from the approach by Duff and Johnson (2001) and substantially goes beyond the latter’s original idea and scope, as will be discussed next.

## 1.2 Ontological Analysis

Duff and Johnson (2001) remain close to the syntactic level of the inquiry when analyzing the given and wanted information entities. The Wantededs remain abstract and unspecific while the Givens describe only isolated information entities with no further meaning or context. Furthermore, the types of questions submitted reveal the essentially archival point of view taken by Duff and Johnson (2001) during the analysis of the inquiries. The Wantededs in particular demonstrate shortcomings in the analysis of and reflection upon the interest of the users.

Since the contents of archival materials cannot be easily conveyed to the user because of the sheer quantities of records available in archives, digital archival information systems and (digital) archival finding aids are eligible means for enabling and facilitating access. These systems rely on the reduction and normalization of the available information pertaining to records, in order to handle formal queries and to retrieve sets of relevant records. Merely knowing the type of given information users bring to an archive and archival information systems, however, does not suffice in order to gain a deeper understanding of user needs and interests, and particularly not for the design of better archival information systems.

The ultimate aim of the study at hand is to contribute to the improvement of automatic query facilities of archival information systems. These automatic queries must be formulated in an ontological form. The intension of the term *ontology* has significantly expanded over the last decade. In philosophy, the term means the study of being. Since the adaption of the term, especially by the computer and information sciences, its scope has grown to include a broad range of data models, controlled vocabularies, or explanatory models. In this study, ontology is understood as a *conceptual model* which comprises formalized knowledge as clearly defined concepts and relationships pertaining to the possible state of affairs in a knowledge domain. Ontology does not denote a data model as in the computer sciences but rather the domain of discourse referenced by the inquiries.

The analysis conducted by Duff and Johnson (2001) exhibits a *semantic mismatch* with a query system. This is because it does not provide the necessary ontological structures with which to investigate how an information system needs to be designed in order to adequately respond to the interest of a real or hypothetical question or to develop new and better data structures. In order to better understand the sense of an inquiry and to identify the subject matter of its interest, the character of the Givens and especially of the Wantededs needs to be further investigated and analyzed on an ontological level.

This study, therefore, further interprets the Wantededs by analyzing the subject matter which typically appears as the *interest* of the inquiry from a user point of view, or, in other words, the common domain of discourse to which the inquiries refer. Duff and Johnson (2001) adhere to the *syntactic form (Aussageform)* of an inquiry while the methodological approach of this study proceeds further to the analysis of the *epistemological form (Erkenntnisform)* of an inquiry. The guiding question during the interpretative analysis asks what the user needs to know in order to satisfy the perceived interest of the inquiry.

In order to answer questions, the first basic prerequisite for an information system is a structuring of its information contents according to an appropriate schema. However, there is a semantic mismatch between the question and the schema because one cannot expect to recognize a schematic entity in the question. An ontology is needed between the schema and the question because the meaning of natural language is only comparable with an ontology. The technical reasons are that our thinking, and hence the meaning of natural language, implies deep subsumptions of hierarchies of concepts and relationships, and hence cannot be represented with relational or XML schemata (Oldman et al., 2014). Examples for ontologies created for natural language processing are CyC<sup>27</sup>, one of the first formal ontologies, and WordNet<sup>28</sup>. No simpler method of representing the meaning of natural language has been found so far.

As has been shown in Doerr (2003) and Doerr and Iorizzo (2008), the representation of relationships is pivotal to such ontologies because queries need relationships as structural parts, whereas terms can be dealt with as data items. For this reason, this study creates an ontology and does not supplement a schema such as EAD<sup>29</sup> because this would render the results dependent on a particular form of the schema, here the tree-like data structure of XML and the tags and attributes of EAD, and would neglect the sense of the questions. The result would only be a projection of the latter, and there would be no control over the adequacy and completeness of the data structure (based on the schema) in relation to the original question. Rather, this study deems it more appropriate to begin with the inquiry, its grammar and words, to proceed to the ontology, and then continue from the ontology to the schema; in doing so, this study focuses on the penultimate step in the process. Different schemas and data structures can be created from the ontology and remain compatible. These data structures can then be queried in an information system. Whether these queries are useful or meaningful can only be determined on the level of the ontology because the latter represents the sense and interest of the inquiries. This relationship between schema and ontology has been discussed, for example, by Gruber (1995, 2007), Guarino (1998), and Calvanese et al. (1998), and has not been challenged since.

More in-depth knowledge of the ontological character of the given and wanted information is therefore needed in order to obtain a deeper understanding of the character and subject matter of the interest of the inquiries. Their interpretation is determined not only by the immediate recognizable sense of the questions and associated contextual information, both made explicit by the given and wanted entities, but also by an epistemological framework. In other words, the inquiries are posed in the context of a particular domain of discourse, which determines how the sense and subject matter of the interest of a question can be further determined through interpretation. This epistemological framework comprises the *archival domain* of record keeping and the *domain of historical inquiry* for which traces and evidence can be expected to be found in the archive. However, the selected ontology for formalizing the results from the

---

<sup>27</sup> <http://www.cyc.com/>

<sup>28</sup> <https://wordnet.princeton.edu/>

<sup>29</sup> <http://www.loc.gov/ead/>

interpretative analysis also influences the interpretative process. Furthermore, the interpretation and ontological analysis of the inquiries rely on educated intuition regarding both domains and necessarily filters probable implicit questions.

The first step in the ontological analysis is the identification and categorization of the given and wanted entities referenced in the inquiries. They constitute the first building blocks of the ontological model.

### 1.2.1 Referenced Entities

While Duff and Johnson (2001) have conducted a predominantly linguistic analysis by identifying different types of information entities as Givens and Wanted, thus adhering to the paradigm of the keyword, this study applies an ontological analysis. In order to create an adequate ontological model, the entities referenced in the domain of discourse of the inquiries need to be identified.

In the case of the *Givens*, for example, Duff and Johnson (2001) categorize according to the type of given information entity such as *proper name* for a given “personal name”, “corporate name”, or any other “name of an entity”. In contrast, this study analyzes which entity is referenced by a Given, for example, a *person* is referenced by a personal name or a *group* is referenced by a corporate name. In order to distinguish between the different conceptualizations, the term *given entity* will henceforth be used instead of “Given”. A *given entity* represents an entity referenced by the user in the question or the associated contextual information by a word or expression in order to describe and qualify the wanted entity.

The given entity is determined by its use in the wider context of the inquiry. For example, “German Democratic Republic (GDR)” may be referred to as a geographical location, as a conscious actor, or as a period denoting the time frame of the existence of the GDR. Furthermore, the given entities are further distinguished into particular entities and general types of entities; for example, a particular person, such as “Konrad Adenauer” or “Erich Honecker”, and a type of person, such as “photographer” or “grandfather”.

In this study, supplementary information is not excluded from coding or analysis. On the one hand, it is difficult to categorize any information as insignificant, especially in an archival setting where contextual information can take on a vital role in discovery processes and the location of archival materials. On the other hand, all information a user provides to describe the information need can be deemed relevant in some way simply because the user decided to provide it. Finally, every item of information provided does carry a potential relevance for the analysis of the subject matter of the interest of the inquiry and should therefore be considered.

In the case of the *Wanted*s, Duff and Johnson (2001) only provide general and highly accumulative categories broadly determining the type of the wanted information such as “form”, “recommendation”, or “general or background”. Determining the Wanted in this way proves to be neither suitable nor productive for the purposes of the study at hand, particularly in view of the unspecific and overly broad categories for the wanted information types determined by Duff and Johnson (2001).

Here, however, the Wantedds will be referred to as *wanted entities*. These represent a type of entity primarily referenced by the wanted archival or non-archival material or by a wanted fact. For example, in the case of a material-finding question for documents about a particular person, the wanted entity would be that particular person. In case of a factual question asking, for instance, for the name of a particular person who has been at a particular place, the wanted entity would again be that particular person.

Since the type of question already indicates whether the user is interested in – or *wants* – a resource or factual information, the wanted entity further substantiates the information the user seeks to obtain by asking the question which is typically not simply the location of a specific document, or general or background information, but relates to a specific type of entity such as a particular person, event, or place. The wanted entity thus also indicates the most recognizable general interest of an inquiry.

Interpreting the Wantedds this way allows our analysis to move away from the immediate utterance of the question and to take a closer look at the actual epistemological interest of the user embodied in the inquiry. To be sure, the Wantedds only approximate the information need and immediately recognizable primary interest of the questions – again beyond an interest in resources or factual information – in terms of the most recognizable referenced entity.

The categories for the referenced given and wanted entities both emerged iteratively from the analysis of the questions and the associated contextual information. Similar kinds of entities were successively allocated broader categories, mainly based on the conceptualization of the ontology CIDOC CRM (IV:3.2), which provides a sound conceptual framework for determining and formalizing the type of referenced entities.

The given entities are grouped into the following seven categories: *Actors*, including persons and groups, *Documents*, *Events and Activities*, *Time*, *Place*, *Things*, and *Other Entities*, including general context and identifiers. The wanted entities comprise as principle categories *Actors*, including persons and groups, *Events*, including unintentional events and activities – both concepts will be further explained later on (IV:3.2) –, *Documents*, *Things*, *Places*, and *General Topics*. Most categories contain one or more sub-categories, further specifying the nature or context of the wanted entity. Both the wanted and given entities are discussed in more detail and exemplified in the next chapter (V:1.2).

The identification and abstraction of the given and wanted entities to common conceptualizations is the first step of the ontological analysis. The occurring phenomena in the inquiries and their domain of discourse are reduced significantly during this process.

However, the ontological analysis of the wanted and given entities is only the initial step away from the level of utterance of the question and towards the identification and ontological formalization of the wider shared domain of discourse and the subject matter of the inquiries.



### 1.2.2 Patterns

The next step in the ontological analysis further interprets the inquiries with regard to their shared domain of discourse. The goal is to ontologically formalize relevant entity types and especially the relationships between these entities, which together describe the shared subject matter of the interests of the inquiries. The previously identified given and wanted entities provide initial building blocks.

Formally, an *ontology engineering* approach is employed in that the inquiries and their interpretations are translated into an ontological model. The domain of discourse described by these entity types and relationships represents the typical and shared subject matter of the interests of the inquiries. The result of this process is an ontology called the *Archival Knowledge Model* (AKM).

As a whole, the AKM can be partially understood in analogy to what has been described as a *frame*. Minsky (1974) has defined frame as a “framework to be adapted to fit reality by changing details as necessary” and as a “data-structure for representing a stereotyped situation” consisting of a “network of nodes and relations”. The AKM resembles such a frame in that it describes a general network of entity types, the “nodes”, and relationships (IV:3.1) between these nodes, and thus also describes general “stereotyped situations” of a past historical reality appearing as the subject matter of user needs and interests of inquiries made of archives.

The AKM can be further differentiated into several *general patterns*, which constitute typical segments or aggregations of the ontology as a whole. These general patterns also resemble frames and consequently embody various specific “stereotyped situations” of historical contexts to which the subject matter of the interest of inquiries epistemologically refer. In other words, they describe typical and specific contexts of user interests communicated to archival information systems.

The AKM and its general patterns are the result of the interpretation and ontological analysis of the inquiries. The main constituents of the ontology AKM are relationships, entity types and scope notes; that is, definitions of the relationships and entity types based on the further analysis of the given and wanted entities, and the relevant relationships that exist between them. Which relationships are relevant depends on the further determination of the common subject matter that adequately describes a historical reality to which the interests of inquiries refer. In other words, entities and relationships together provide the necessary historical and contextual knowledge in order to formulate a query that would retrieve potentially adequate documents or facts serving the perceived interest of an inquiry.

Since the inquiries are not self-contained, the determination of said subject matter is further based on the wider epistemological context of the interpretation, which includes the historical and archival domain (IV:2), as well as common sense background knowledge. Both domains provide concepts necessary to adequately represent the subject matter of the interest of an inquiry.

A second principal type of pattern which primarily serves demonstrative purposes will be

introduced at this point. The general patterns are situated on the schema level of the AKM and hold entity types and relationships as the general concepts of the relevant phenomena which exist in the domain of discourse. The instance level contains particular examples of the general entity types and relationships, and together with the schema level constitutes the knowledge base. Here we will distinguish between two types of *instantiated patterns*.

The first type of instantiated pattern is that of *query patterns*. The general patterns can be instantiated with specific examples of entity types and thus adapted to the aforementioned “specific needs”. In the context of this study, these “specific needs” are the specific interests of inquiries. Query patterns represent instantiated general patterns with given entities from an inquiry and specify a primary query target. In other words, query patterns indicate potential adequate queries for retrieving relevant documents or facts in order to respond to the interest of a question.

As will be discussed in greater detail in the next chapter (V:1.5), query patterns always instantiate elements from two general patterns: one provides entities which describe the provenance of the wanted archival materials or facts while the second provides entities which represent the “aboutness” of the same.

The query patterns are based on the assumption that machine-supported search and discovery systems may effectively process data described by a schema based on the AKM. The search system; that is, the query patterns, allow potentially unlimited numbers of queries as instantiations of general patterns and thus a virtually unlimited number of questions.

As will be discussed in more detail during the ontological analysis of the inquiries (V:1.3), the interpretation of questions often allows for conclusions on more than one potential adequate query, so-called *indirections*. Since distinct query patterns cannot be created for each individual potential query, query patterns typically subsume more than one. In this regard, the query patterns neither constitute efficient query formulas nor do they even cover every potential indirection. The query patterns illustrate that the interest of a question can be served by a formal potential query against the subject matter described by the AKM. As such, the AKM provides a relatively small set of relevant entity types and relationships necessary for creating data structures adequate for serving typical interests of archival inquiries.

The second type of instantiated pattern is called an *archival pattern* and is solely used in the chapter “Application” (VI). While query patterns indicate potential queries and exemplify the relationship between the inquiries and the ontology, the AKM, archival patterns show how real-life metadata, here encoded with EAD (VI:1), can be represented in general patterns of the AKM. In this regard, archival patterns exemplify the relationship between schemas and data structures and the AKM. Again, archival patterns are auxiliary vehicles for demonstrating how real-life data can be analyzed and that the latter already contain explicit and implicit information relevant to the AKM.

The general patterns, query patterns, and archival patterns will be visualized in diagrams. The notation used in these diagrams will be introduced later in this chapter (IV:3.3).

### 1.3 Summary

The interpretative analysis of the inquiries is conducted on two different yet closely connected levels. The first step is a *linguistic analysis* – understood in the sense of studying the semantic level of immediate meaning of words in a phrase – where the inquiries are categorized according to whether archival or non-archival materials, factual information, or other services are requested. This categorization allows for an initial appraisal of the interest of the inquiries and a separation of substantial inquiries – resource-discovery and fact-finding – from non-substantial inquiries – non-discovery – for further analysis.

While the linguistic analysis focuses on the level of utterance, the *ontological analysis* further interprets the inquiries in order to better understand their sense. The goal here is to identify shared interests and to formalize and explicate the subject matter of these interests to common and abstract ontological conceptualizations. The result of the ontological analysis is an ontology, the AKM, which contains entity types and relationships along with scope notes describing and representing an adequate historical reality as the common domain of discourse of the inquiries.

The AKM is comprised of *general patterns* focusing on typical segments of the historical reality to which the interest of inquiries refers. General patterns thus represent shared and principal interests of inquiries in abstract and formalized ontological terms. *Query patterns* represent an instantiated unification of two general patterns and demonstrate the relationship between an inquiry and the AKM by exemplifying how a specific interest could be formulated as a categorical query using selected elements from general patterns. *Archival patterns* then show how real-life data can be represented with the AKM.

In order to identify and define general patterns, the *ontological analysis* follows the linguistic analysis and is conducted in two iterative steps. The first step investigates the given and wanted entities as previously discussed. The given entities refer to ontological entities given in the inquiries in order to qualify and describe what the user wants. The wanted entities indicate the general interest of the inquiry and refer to the most recognizable and primary ontological entity in which the user is interested. Both the given and wanted entities provide the basic building blocks for the general patterns. However, the determination of the type of question and analysis of the given and wanted entities remain focused on the immediate recognizable sense of the inquiry and do not suffice for an adequate ontological representation of the subject matter of the shared interests.

The second step in the ontological analysis therefore subjects the wanted entity in the wider context of the inquiry to a more in-depth analysis. The goal is a better and deeper understanding of the sense and subject of the interest of the inquiries, the identification of the shared and common conceptualizations, and their formal representation in an ontology. This in-depth interpretation goes hand in hand with the translation from the utterance level of the inquiries to the ontological level of the general patterns, which are its primary outcome.

During these steps, the interpretative analysis will adhere as far as possible to the requirements of the interests as expressed in the inquiries. Accordingly, the ontological modelling

will represent foremost entities which are sufficiently backed by the empirical evidence found in the inquiries. Additional context will be included only if deemed essential to represent an adequate answer in the context of general patterns or to enable important secondary or auxiliary questions.

This second step of the ontological analysis will be the main subject of the following sections. Here we will focus in detail on the epistemic context of the interpretation of the inquiries, and on the process and means with which the results from the interpretative analysis can be formalized and explicated via ontological modelling.

## 2 Epistemological Framework

The interpretation of the single questions and associated contextual information as well as their further ontological analysis are necessarily determined and guided by a wider epistemological framework. *Epistemology* here refers to the fundamental question applied during the interpretation of what and how the user seeks to learn via the inquiry. The epistemological framework evolves foremost from the recognizable interests found in the inquiries themselves. The given and wanted entities constitute the initial building blocks for the resulting ontological representation of the domain of discourse, as they describe what the user knows and wishes to know.

The *historical domain* represents the domain of discourse assumed to provide the framework for the principle epistemological interest of the inquiries; that is, questions about past human activities in the widest sense. The scope of the interpretative analysis is thus limited to particular kinds of questions as well as answers pertinent to the historical domain, and excluding other knowledge domains such as archaeology, philosophy, physics or biology questions and their possible interests. While the *archival domain* evidently constitutes the domain of discourse prompted to provide expedient knowledge with which to answer inquiries, it may of course itself be the object of inquiry. With regard to the historical domain, the *background knowledge* of the interpreter<sup>30</sup> further implicitly determines the epistemological framework. This background knowledge has been incorporated in the following discussion about the historical domain. Finally, *common knowledge* about the workings of the world constitutes another essential source of a vast range of highly implicit knowledge.

The ontology AKM pertains to the subject matter of the historical and archival discourse as expressed in the inquiries. The understanding of the epistemology of the users posing these inquiries and of the underlying historical and archival discourse provides criteria with which to further decide on the relevance of the kinds of facts from this subject matter that deserve modelling in this ontology; in other words, the kinds of facts which appear of primary and common interest to the users and the underlying archival and historical discourse.

---

<sup>30</sup> In this case, the author of this study as a trained historian has background knowledge about the principle interests of historians and relevant contexts for historical inquiry.

It is important to bear in mind that the objective of this study is not to incorporate or translate theoretical concepts or the perceived epistemology of one or both knowledge domains into an ontological model. The immediate goal is not to investigate the appropriateness of archival documentation and description principles in relation to the perceived user needs. On the contrary, the aim here is to firstly lay the groundwork for such investigation, based on an empirical foundation, the inquiries, and to extrapolate appropriate ontological formalizations.

The epistemological framework delineated in the following section provides a fundamental context of the epistemology identified in the inquiries: A historical question strives to obtain specific types of historical knowledge, and questions posed to an archive may only expect particular kinds of answers. In short, the epistemological framework determines which kinds of explicit and implicit questions and answers are probable as well as adequate in the context of their interpretative analysis.

## 2.1 Historical Domain

As signaled in the previous chapter “Empirical Data” (III), the collected inquiries arise from the historical domain, in the widest sense understood as the histories of the human past. Even though the information about the inquiring users is scarce, the historical domain and its methods can provide guidance and a basic framework for the interpretative and ontological analysis of the questions. The following section delineates essential characteristics of the historical domain in terms of common and principal research interests and methods. The account is based on a selection of texts on history as a scientific discipline, and is also inevitably influenced by the background knowledge of the author as a qualified historian.

**History** The term *history* carries three discernible senses: the term may denote (1) what has happened and transpired in the past, (2) the study, presentation, and narration of that past, and (3) a particular form of knowledge about said past created based on particular methodological regulations (Boshof et al., 1997, 2). These three senses are, however, intrinsically intertwined and their essence has been subject to extensive debate in the field of philosophy of history ever since historians have thought and written about the past (Boshof et al., 1997, 2).

The *domain of discourse* of history comprises all that has happened in space and time and which is tangent to human life. In the words of Bloch and Le Goff (2008), the good historian resembles the man-eater from fairy tales who finds his prey where he senses human flesh.<sup>31</sup> The principle and natural *subject matter* of historical inquiry is therefore essentially an *anthropological* one since humans “make” history; that is from the characteristics of landscapes, create tools and machines, compose documents, or establish and run institutions (Bloch and Le Goff, 2008, 30). Historians seek to describe and understand the history pertaining to human activity (*Tun* or *Handeln*) – what humans have done – and human suffering (*Leiden*) – what has happened to

---

<sup>31</sup> “[Der] gute Historiker [gleich] dem Menschenfresser im Märchen. Seine Beute weiß er dort, wo er Menschfleisch wittert” (Bloch and Le Goff, 2008, 30).

humans – in the past, either as individuals or in the context of social groups (Faber, 1982, 35). The term “suffering”, in this context, not only refers to concrete physical and psychological afflictions but also comprises all external influences of “historical forces” and imprints experienced by humans and which they cannot directly or consciously control or influence, such as climate, vegetation, economy, society, or religion, all of which have rendered humans “suffering” objects throughout history (Boshof et al., 1997, 3). History is “for” the self-knowledge (Collingwood, 1970) of humankind, its self-certitude<sup>32</sup>, and the attempt to provide accountability for its own past.

Boshof et al. have stated that this principle domain of discourse constitutes a *regulatory idea* for historians undertaking historical inquiry. Naturally, historical inquiry can always only focus on particular fragments of the past. The historian attempts to render selected fragments of the past current by studying, presenting, and narrating them to contemporaries. This endeavour is necessarily subjective, determined by the interests, questions, and intentions of the inquirer, the historian, and of what the current era, the contemporaries, find interesting about the past epoch (Boshof et al., 1997, 3). The resulting historical knowledge, the “history”, represents neither “the past”, nor yet the surviving past” but “a reconstruction of certain parts of the past (from surviving evidence) which in some way have had relevance for the present circumstances of the historian who reconstructed them” (cit. Tosh, 1996, 145-146).

According to Boshof et al., the natural sciences (*Naturwissenschaft*) seek to describe and explain the workings of the world based on general laws established based on (repeatable) experiments and expressed as generally valid propositions, history as a science focuses on the singular and unique, and aims at the accurate temporal fixation of their causes, courses, and ramifications. This focus on the singular and unique, causation, historico-genetic perspective, understanding (*Verstehen*), and temporal anchorage are key criteria of the historical method (Boshof et al., 1997, 9). The terms *idiographic* (describing the singular) and *nomothetic* (establishing laws), which have been coined by Windelband (1904), express well the fundamental difference in these two approaches to (scientific) knowledge.

To put it another way, the past cannot be recreated in experiments and the historian cannot deal directly with past events but only with a “statement about the event (...) which affirms the fact that the event occurred” (Becker, 1955, 330). Historical research is permeated by a fundamental “distinction between the ephemeral event which disappears, and the affirmation about the event which persists” (Becker, 1955, 330). The historian may find such affirmations substantially within the archive in symbolic and mostly written form as “histories” about past events. The empirical foundation for historical inquiry into the human past essentially constitutes the surviving remains of past human activity within an archive. Dealing with affirmations carried by these remains, connecting and making sense of them, ascertaining whether they are true or not, accurate or not, is the business of historians. Their *scientific* treatment is regulated by the historical method.

---

<sup>32</sup> “Die Geschichte ist das Wissen der Menschheit von sich, ihre Selbstgewissheit” (Droysen, 1868, 38).

**Historical Method** According to Rüsen et al., the *historical method* is understood as the embodiment of those regulations which determine historical thinking as a cognitive process and which substantiate the claim of truth in historical writing. Traditionally, these regulations of historical thinking have been found and expressed in two different kinds of operation beginning with the dominance of the literary form of historical knowledge and then, since the 19th century, the scientific production of historical knowledge through research (Rüsen and Jaeger, 1990, 13).

This study restricts its scope to the scientific form of historical knowledge production. As Rüsen et al. continue, the historical cognitive process consists of the three basic operations *heuristics*, *critique*, and *synthesis* – also called interpretation – which are determined by common and authoritative regulations underpinning and guiding their application. These three basic operations are evident in the various historical research processes. The reconstruction of the human past from empirical evidence is thus organized into a coherent principal methodological structure. The historical method as a systematic unity of heuristic, critique, and synthesis has been first theoretically described and practically employed during the *Historismus* era, and is primarily connected to Droysen (1868) who established and propagated this tripartite working mode as historical method. The latter provides a principal methodological framework for the scientific conduct of historical inquiry (Rüsen and Jaeger, 1990, 13-18).

In the historical sciences, *heuristics* – the “art of finding” – constitutes the first principle step which denotes the process of *finding* a relevant historical research question and, at the same time and most importantly for this study, of *finding* adequate sources of evidence and factual information pertaining to the research question (Rüsen and Jaeger, 1990, 13-15).

The *historical sources* (*Quellen*) constitute the empirical basis for historical inquiry. As a diverse and vast body, historical sources may comprise any kind of “human remains and such products of man’s activity as either were meant by their authors to communicate knowledge of historical facts or conditions, or by their nature are calculated to do so” (Shafer and Bennett, 1980, 103). This idea of intentional and unintentional sources had been distinctively formulated by Bernheim (1903) who distinguished the principle character of historical sources into *traces* (*Überreste*) and *traditions* (*Traditionen*). Historical sources have the character of traces if they have been created without the intention or purpose of future remembrance of the past such as orders or crew lists, but also oral, abstract and material entities such as facts pertaining to buildings or society. Sources have the character of tradition if they have been intentionally created with the purpose of being a testimonial and to serve remembrance of the past such as biographies or autobiographies, memoirs, accounts of journeys or chronicles (Bernheim, 1903, 230-234).

The distinction of Bernheim (1903) is not absolute but functional, as are other categorizations of historical sources such as according to primary and secondary nature. Any historical source may be interpreted either as a trace or tradition, often depending on the particular historical inquiry. Historiographical traditions may include recent “secondary” scientific literature such as journal articles or monographs, especially if these works are the object of inquiry; however, mostly the term refers to contemporary “primary” narratives characterized by personal and

subjective accounts of experiences, events or activities. In the end, any source has factual and evidential potential and, at the same time, carries a sense determined by those historical subjects who created them for particular reasons. The facts and propositions found in either traces or traditions constitute empirical facts and are historically relevant: Why does someone claim or state a fact, regardless of whether the fact is true or not? And is the fact indeed true?

Regarding the substance (*Überlieferungssubstanz*) of historical sources, written documentary traces constitute the quantitatively largest and for most kinds of inquiries qualitatively fundamental type of historical sources, in particular various kinds of documents (*Geschäftsschriftgut*) of corporate bodies such as *official documents* (*Urkunden*) and *files* (*Akten*) (Boshof et al., 1997, 216-226).

According to Boshof et al., *official documents* are attested and created according to certain formal requirements pertaining to acts of a legal nature. Examples of official documents include edicts, notata, charta, laws or decrees as well as orders. An important task and research inquiry for historians pertains to the investigation of the extent and form of the actual execution of legal acts, and the discovery of proof of its factual execution (Boshof et al., 1997, 228-230). The largest group of traces form the written *files* (*Akten*) which particularly pertain to phases of preparatory deliberation and expression of will in the form of memoranda and drafts as the result of the work of corporate bodies. Furthermore, many additional pre-forms of official documents and working papers are found in files which may also have legal status. Files are complemented by many other kinds of documents pertaining to the course of business (*Geschäftsgang*) of corporate bodies; for example, plans for office organization (*Büroorganisationspläne*) or file distribution (*Aktenverteilungspläne*), papers and documents which have moved between organizational units of corporate bodies (*Schriftstücke der Überordnung, Unterordnung, Gleichordnung*), as well as notes, minutes, memoranda, or other kinds of written records (*Aktennotiz, Protokoll, Aufzeichnung, und Denkschrift als neutrales Schriftgut*) (Boshof et al., 1997, 230-246).

Other kinds of written traces, according to Boshof et al., are journalistic sources about the transmission of news and opinions in the widest sense, letters, diaries, or other kinds of records of conversation, as well as literary traces such as fiction writing and poetry, unpublished personal estates, or specialized literature. Non-written traces may include institutional, linguistic, or social facts of the present such as particular forms of marriage, languages, or professions, physical things such as commodities, works of art, buildings, or even human corpses, visual remnants such as paintings, photographs, or films, and, lastly, audio documents (Boshof et al., 1997, 12-13).

In the context of heuristics, the archive provides the means to discover and obtain potentially relevant historical sources which may contribute factual or evidential pieces towards responding to a historical research question.<sup>33</sup> As Tosh (1996, 65-66) has pointed out, "(...) historical

---

<sup>33</sup> Editions of historical primary sources are, in the end, always products of their time and constitute only particular selections, often displaced from the original archival holding. Archives and their holdings are therefore not merely auxiliary infrastructure but indispensable prerequisites for historical research.



research is not a matter of identifying *the* authoritative source and then exploiting it for all it is worth, for the majority of sources are in some way inaccurate, incomplete or tainted by prejudice and self-interest. The procedure is rather to amass as many pieces of evidence as possible from a wide range of sources – preferably from *all* the sources which have a bearing on the problem in hand”. The heuristic process may further consult other kinds materials such as recent scientific secondary literature from libraries, or primary and secondary accounts of contemporary witnesses.

The second step, the *source criticism* (*Quellenkritik*), denotes the process of applying several principle questions to the inner and outer form of the collected historical sources (Rüsen and Jaeger, 1990, 15-16). The purpose of this step is to determine and establish the reliability, credibility and authenticity of the traces and traditions as evidence or sources of information. Garraghan (1957, 168) delineates several principal questions which target the outer and inner form of the source by asking when, where, and by whom the source has been created, which pre-existing material the source utilizes, and in which original form the source has been produced, and, with respect to the inner form, assess the evidential value of the contents of the source, in other words its credibility.

The third step, *synthesis*, also called interpretation, denotes the application of historical reasoning to the facts, which have been established by inner and outer source criticism, and puts these facts into chronology and context, thus composing a history (Rüsen and Jaeger, 1990, 17-18).

According to Rüsen et al., every historical question entails the conception of an evolving timeline comprised of single facts. This conception is based on conjecture and assumption and thus constitutes the idea of a *possible* history. Historical science transforms this idea of a possible history into a “real” history. Source criticism provides the building blocks for this historical reality, and synthesis assembles and layers them according to particular construction plans for this historical reality. Such construction plans are conceptions of comprehensive timelines into which the facts can be inserted or fit according to their actuality. The connections and coherence created by interpretation sprout in a state of narration; that is, through the execution of historical narration as the constitutive condition of historical knowledge (Rüsen and Jaeger, 1990, 17).

There has been discussion within the field of philosophy of history as to whether or not there is such a thing as patterns of interpretation (Christianson, 1991). Indeed, the key focus of this study rests on the first step of the delineated historical method: the finding and discovering of relevant historical sources. Their interpretation and, in consequence, the assessment of their actual value and use to a particular research endeavour, remains the responsibility of the historian.

This principal three-step approach of the historical method reflects the will to be as objective and accurate as possible, in the words of Leopold von Ranke: “[zu] zeigen, wie es eigentlich gewesen” (Ranke, 1824, V-VI).

**Historical Research** The scientific investigation of the principle domain of discourse follows the historical method and is further based on general methodological principals guiding which and how aspects should be collected from sources in order to reconstruct a history of the human past. It can be differentiated, according to Rüsen and Jaeger (1990, 18), into the two principal categories *hermeneutics* (*Hermeneutik*), and *analytics* (*Analytik*). The differentiation is neither exclusive nor selective; however, it provides further indications towards basic and essential interests of historical inquiry.

Rüsen et al. characterize the *hermeneutic* research strategies as focusing on the sources themselves as potential carriers of the reasons and intentions guiding past human activity and of the self-conception of the actors. In other words: Sources are selected based on preceding heuristic suppositions which give preference to traditions allowing for a reconstruction of past activities (*Handeln*) and suffering (*Leiden*) by understanding (*Verstehen*) the intentions and reasons seen as inherent in these sources. Preferred sources are, for example, documents of political acts such as administrative files or official documents, contemporary historical accounts, or personal testimonials of actors such as autobiographies. History is primarily to be understood and reconstructed from the sources which, of course, do not speak for themselves but contain a kind of pre-narrative (*Vor-Erzählung*) providing an account of changes and activities from one perspective of historical experience. Historical knowledge then is generated from *understanding* (*Verstehen*) of the traditions imprinted in the sources. Hermeneutic research strategies have been formative for the era of Historismus and remain defining for historical method to date (Rüsen and Jaeger, 1990, 18-20).

According to Droysen, the essence of the historical method is to understand through research; that is, by interpretation.<sup>34</sup> Still today, as Rüsen et al. state, historical processes and past events are reconstructed from activities and explained by the intentions, motives and reasons – the inner psychological conditions of actors conveyed indirectly by the sources – that have determined these activities within the interaction of actors. The focus rests on facts (*Tatsachen*) that are understandable; that is, expressions, manifestations, symptoms, or indicators of “meaningful intentionality” (*sinnhafter Intentionalität*) of human life and interaction. Interpretation connects these facts to temporal processes which make sense or have an “understandable” sense (Rüsen and Jaeger, 1990, 18-21).

Historical phenomena and processes should not only be described and understood, according to Boshof et al., but also analyzed and explained. Theories and models are thus introduced, applied to the sources from an external perspective. They introduce measurements which are not generated from the sources themselves. For example, the *Idealtypus* by Max Weber seeks to describe typical attributes of a phenomenon not simply based on empirical evidence but as an illustrative abstraction, provoking tension between empirical essence and conceptual cumulation and abstraction (Boshof et al., 1997, 10-11).

Such *analytical* research strategies, according to Rüsen et al., do not primarily rely on the

---

<sup>34</sup> “Das Wesen der historischen Methode ist forschend zu verstehn” (Droysen, 1868, 9).

potential sense imprinted in sources, the intentions and reasons guiding human activities and self-understanding of actors, but on the external conditions and abstract factors which influence and determine human life and activity; for example, climate, geology, food production, price development, classes and their possibilities, abstract systems which determine what happens and less the actors themselves. Analysis asks first and foremost for facts (*Tatsachen*), which can be subsumed under abstract regularities such as birth numbers, increased numbers of production, and may even take on a statistical, quantifying form. Analysis renders facts fit for theory or uses theory and models for interpretation such modernization theory, fascism or imperialism theory. It is only within such constructs that facts gain historical relevance and meaning (Rüsen and Jaeger, 1990, 20-21).

Historical inquiry conducted within the boundaries of the aforementioned regulative idea of an essentially anthropological domain of discourse and subject matter needs to account for both the structural and intentional determinants of human life, and to consult the historical sources for information and evidence. Tosh (1996, 142) has summarized the essence of historical inquiry as “*selection* – of ‘relevant’ sources, of ‘historical’ facts and of ‘significant’ interpretations”. For the *selection* primarily of ‘relevant’ sources, and to a lesser degree of facts and interpretations understood as secondary interpretative accounts, either contemporary of the past or recent of the present, the archive remains primary and indispensable for the historian.

The presented instruments do not form a definite set; rather, historical inquiry is an evolving process which utilizes theories and methods as set pieces. Their applicants shape and form them while scientifically engaging in historical inquiry and striving for historical knowledge (Boshof et al., 1997, 1-2). However, history as an empirical science necessarily turns to the archive in order to find and discover surviving traces and traditions as the remains of past human activity, which may contain pieces of information or evidence potentially key to a wider research question.

## 2.2 Archival Domain

The inquiries collected arise from the domain of historical inquiry but at the same time also pertain to an archival context so that the epistemological framework of the interpretative analysis also extends into the archival domain and involves its essential principles and constituents.

Since the inquiries are submitted to archives, the principal types of materials and information provided as well as typical services and means of access offered by archives determine the possible expectations applied to the interpretation of archival inquiries. For example, archives do not normally offer extensive bibliographical reference services beyond specialized library sections but first and foremost enable access to archival materials as primary sources. On the other hand, the interest of inquiries may also focus on aspects which can be considered genuinely archival such as the provenance or origination of archival materials.

Archival traditions and practices vary in different countries and even between archives. There are, however, basic principles to which nearly all archival institutions adhere. Several

international archival standards and reference models have been created during the last few decades, substantially advancing efforts towards national and international standardization and interoperability within the archival community. These standards and models formulate common and basic concepts for various areas of archival documentation. The *Encoded Archival Description* (EAD) (VI:1) data model for encoding archival finding aids, for example, is based on the *General International Standard Archival Description* (ISAD(G)) (International Council of Archives, 2000). Other examples include the *International Standard for Describing Institutions with Archival Holdings* (ISDIAH) (International Council of Archives, 2008), the *International Standard Archival Authority Record for Corporate Bodies, Persons and Families* (ISAAR (CPF)) (International Council of Archives, 2004), or the *International Standard for Describing Functions* (ISDF) (International Council of Archives, 2007), all of which target specific aspects of archival documentation and descriptions of archives and their functions.

The following necessarily simplifying discussion of the archival domain will provide an overview of its commonly shared principles, which are essential to this study and the interpretative analysis applied to the inquiries.

**Archives** The word *archive*, according to Franz, can be traced back to the Greek word *arché* which means public authority or public office (*Behörde* or *Amtsstelle*). The primal function of archives has been to store, organize and preserve the mostly written documents and records of governmental and administrative agencies, the *scripturae publicae*, for legal and administrative purposes. It was only later that archives became the memory repositories of society and evidential reservoirs for historians as they are mostly known today (Franz, 2004, 13).

According to the traditional schema by Franz (2004), archives can generally be differentiated into *state archives, communal archives, family archives, ecclesiastical archives, business archives, parliamentary, union, and party archives, archives for literature, arts, and sciences, and press, broadcasting, and film archives*. As the labels for the different kinds of archives already indicate, one of the main differences lies in the scope of the preserved archival materials.

Furthermore, based on Franz, these different types of archives may be further grouped into *public* and *non-public* or private archives. *Public* archives are developed systematically according to particular legal and administrative responsibilities and in the context of the public and administrative bodies they serve. All *non-public* or private archives such as family archives, ecclesiastical archives, or business archives derive from the archetype of the public state archive and are typically organized in much the same or in a similar way. Archival materials may also be preserved in non-archival institutions such as libraries, museums, or scientific institutions, or are housed within private collections or personal estates (Franz, 2004, 19-26).

As already discussed in the previous chapter “Empirical Data” (III), the inquiries under study were directed towards public *state archives*, which therefore constitute the primary focus of the following discussion.

**Archival Holdings** Archives collect and preserve *archival materials*. Following Gilliland-Swetland (2001, 223), the term connotes the “official records, personal papers and manuscripts, audio and visual materials, and realia” commonly found in archives. These archival materials are the *records* of corporate bodies, families and individuals and have been created or otherwise accumulated as the organic by-products of their past activities, typically carried out in the context of particular functions or responsibilities. These mostly unique materials become archival when they are “assembled and preserved as evidence of that activity” (Fox et al., 1998, 5). As such, archives and the records they hold provide legal and primary historical evidence of past activities of their creators and thus also serve as secondary sources of historical information (Fox et al., 1998, 5).

Even though archival materials may be of any form, written materials constitute by far the largest group of records not only in state archives. The customary classification of written *official records* distinguishes *administrative files* (*Verwaltungsakten*), *official documents* or *certificates* (*Urkunden*), and *registers* (*Amtsbücher*) (Franz, 2004, 43-58). State archives in particular are responsible for the collection and preservation of such records of national and state governments and their various agencies.

*Administrative files*, following Franz, are aggregations of a variety of items created or accumulated by a corporate body while carrying out its functions and organized according to internal organizational systems or other specific requirements. Such organizational systems may include, for example, simple arrangements of records according to chronological aspects in series (*Serienakten*) and sub-series (*Teilserien*), or more systematically pertaining to general cases (*Vorgänge*) or specific subject areas (*Sach- und Betreffakten*), which may be differentiated further into more specific subject areas (Franz, 2004, 52-54). Such administrative files would then typically contain those materials that were created and accumulated until the conclusion or settlement of a particular case (Franz, 2004, 35), such as decisions and plans made, background information obtained, statements formulated and communicated, or notes and instructions passed along departments. As such, administrative files in particular illustrate the inner working processes of corporate bodies and other groups.

Further, *official documents* are related to a legal act (*Rechtsakt*) which, in recent times, has corresponded to the act of administering an official document such as private contracts or state treaties, appointments of public officials, or judicial decisions (Brenner-Wilczek et al., 2006, 49-50). These documents may be part of administrative files.

*Registers* pre-date administrative files and typically represent hardback pages or books containing entries serving a broad range of different purposes regarding the daily business of a corporate body (Franz, 2004, 55-57). For example, registers have been used to record short summaries or copies of official documents or letters, minutes of meetings and decisions taken, as well as inventories of any kind, for accounting, or personnel lists (Franz, 2004, 55-57).

*Personal estates* represent an important part of archival holdings. Personal estates may contain any kind of archival materials but also hold items that are typically not found in other records

such as private letters and correspondence, diaries or notebooks, manuscripts and drafts (Franz, 2004, 67-68), but also various kinds of realia.

Furthermore, archives often also maintain different kinds of *special collections* as complements to archival records such as collections of newspaper cuttings, posters or leaflets, as well as collections of maps and other kinds of geographical or topographical plans which are often held in special departments separate from the original files. More recently, photographs and audio documents, and especially electronic data carriers constitute another special type of collection often held in specialized archives (Franz, 2004, 60-71).

Finally, archives also hold and preserve printed *publications (Druckschriften)*, created or collected by institutions as part of their activities and functions, for example, parliamentary publications, budget plans, annual reports, statistical or informational publications, or other individual publications (Franz, 2004, 58-59).

All these different kinds of archival materials constitute the holdings of an archive. However, in contrast to the materials collected by libraries such as published books or journals, most archival materials are individual and unique items of which no multiple copies exist. Furthermore, materials originating from the context of the same activity typically constitute a complex body of “organic inter-relationships” – particularly illustrative in the case of administrative files – from which single items derive their particular meaning and authenticity (Gilliland-Swetland, 2001, 201-202). The identifiable and basic unit of interest in the archival context is therefore typically an aggregate of archival materials such as files or series, but not a single item (Pitti and Duff, 2001, 1-2).

These archival specifics have ramifications on how archives arrange and describe their holdings. While special collections and printed publications in archives are typically arranged and described by criteria unspecific to the archival context such as by content criteria – classification or keywords (Brenner-Wilczek et al., 2006, 56-57) –, or by creator (Franz, 2004, 58-59), the arrangement and description of official records and personal estates adhere to the particular archival principles of the *respect des fonds*.

**Archival Principles** The official records, and in most cases personal estates, are typically organized into *fonds*, which represent aggregations of often hundreds or even thousands of single items. Fonds are arranged and described according to the principle of *respect des fonds*. The principle states that “the origins of the assembled materials as an integral and organic corpus of documentation” (Fox et al., 1998, 6) must be carefully reflected when arranging and describing records in fonds by respecting their *provenance* and *original order*.

Both principles guide *archival arrangement* understood as the “intellectual and physical processes and results of analyzing and organizing documents in accordance with archival principles” (International Council of Archives, 2000, 10). In this context, the *principle of provenance* states that all records in a fonds should share the same provenance, meaning that they originate from

the same corporate body, family or individual (Pitti and Duff, 2001, 1-2).<sup>35</sup> The *principle of original order* then states that the original inner structure of the records in a fonds should be preserved (Haworth, 2001, 12) and that records of different provenance should not be mixed. Only in cases where no pre-existing internal structure is apparent may the archivist create a rational order sensitive to the nature and use of the archival collection (Fox et al., 1998, 6-7).

While the principle of respect des fonds was originally developed based on experiences with official records of state and governmental provenance, the principle also applies to the records of families and individuals (Fox et al., 1998, 6), which are treated as logical units if the contents of the records are “intrinsically bound up with the life of the individual or the functions of the organization from which they emanated, and cannot be fully understood apart from them” (Fox et al., 1998, 6).

In contrast, primarily during the 19th century and following the older *principle of pertinence*, as Franz points out, archival materials were assembled and arranged according to subject matter and regardless of their origin and context of creation; for example, according to the geographical places to which documents refer. Even though the arrangement according to primary subject matter facilitated specific questions; for example, concerning the history of particular geographical places, other as yet unknown questions and research uses of an archive are difficult to anticipate for the creator of the archive or the archivist seeking to describe it. Furthermore, arrangement according to pertinence becomes intractable as soon as archival materials refer to more than one subject matter (Franz, 2004, 34-35).

Even though fonds “inherently reflect the biases and perspectives of their creators”, the records may provide “authentic, accurate, and impartial evidence” of the activities of their creators if they have been arranged and described according to the respect des fonds principle (Gilliland-Swetland, 2001, 202). Only then can the forces, activities, and functions that produced the records be understood, the contents of archival materials interpreted, and their authenticity and reliability as evidence of past activities assessed (Fox et al., 1998, 6). Finally, the sheer quantities of archival materials necessitate descriptions of the “bigger picture”. Today, therefore, the principle of respect des fonds is the established and applied principle for the arrangement of archival materials.

Accordingly, *archival description*, understood as the “creation of an accurate representation of a unit of description and its component parts, if any, by capturing, analyzing, organizing and recording information that serves to identify, manage, locate and explain archival materials and the context and records systems which produced it” (International Council of Archives, 2000, 10), adheres to the principle respect des fonds. In other words, archival description conveys information about the *internal* structure (original order) and the *external* structure (provenance and origination) of a fonds (Haworth, 2001, 12).

Fonds of unique and interconnected archival materials “ranging in size from several to

---

<sup>35</sup> “Originator” does not necessarily imply “creator” but only the actor who turned the archival materials over to the archive.

several million individual items” are very different from the printed monographs or serial publications collected and described by libraries (Fox, 2001, 63). Bibliographical descriptions represent “individual published item[s]” and therefore exhibit “a one-to-one correspondence” to the described item (Fox, 2001, 62-63). The contents of the bibliographical description are transcribed from the published work with the underlying assumption that “such data captures the information needed for finding and identifying” relevant items and that the subject matter of the published work adequately represents user interests (Fox, 2001, 62-63). In contrast, archival description represents a fonds as “the output of an organic activity” (Fox, 2001, 63) of an actor where the description corresponds to different levels of aggregates of unique items, all of which share the same provenance (Pitti and Duff, 2001, 1-2) and lack “the inherent identifying characteristics of published works” (Fox, 2001, 63). Therefore, “archival materials can only be found and identified on the basis of their origins, expressed by identifying both their creator and the functions that they document” (Fox, 2001, 63).

Furthermore, archival descriptions strive to reflect and describe the records’ original arrangement; that is, how a record-creator has organized records (Haworth, 2001, 14). Apart from descriptive fields providing respective information, the arrangement of the fonds according to its original order is reflected in a *multilevel description* involving “a complex hierarchical and progressive analysis” (Pitti and Duff, 2001, 1-2). This starts from the description of the whole fonds and then subsequently identifies and describes significant sub-parts in a multilevel hierarchy where each level is linked to its immediate next-higher predecessor (Haworth, 2001, 14-15). The multilevel hierarchy normally reflects the “intellectual structure and content of the material” but not necessarily the exact physical characteristics (Pitti and Duff, 2001, 1-2). The archival materials become intelligible from their context within this multilevel description.

The exact number and scope of the different levels included in such multilevel descriptions may vary with regard to the provenance and characteristics of the archival materials as well as according to national traditions.<sup>36</sup> Figure 3 depicts a possible hierarchical model including some of the most common and widely accepted types of levels (International Council of Archives, 2000, 36).

---

<sup>36</sup> In the context of this study, the exact levels are of no further importance but are briefly discussed for illustrative purposes.



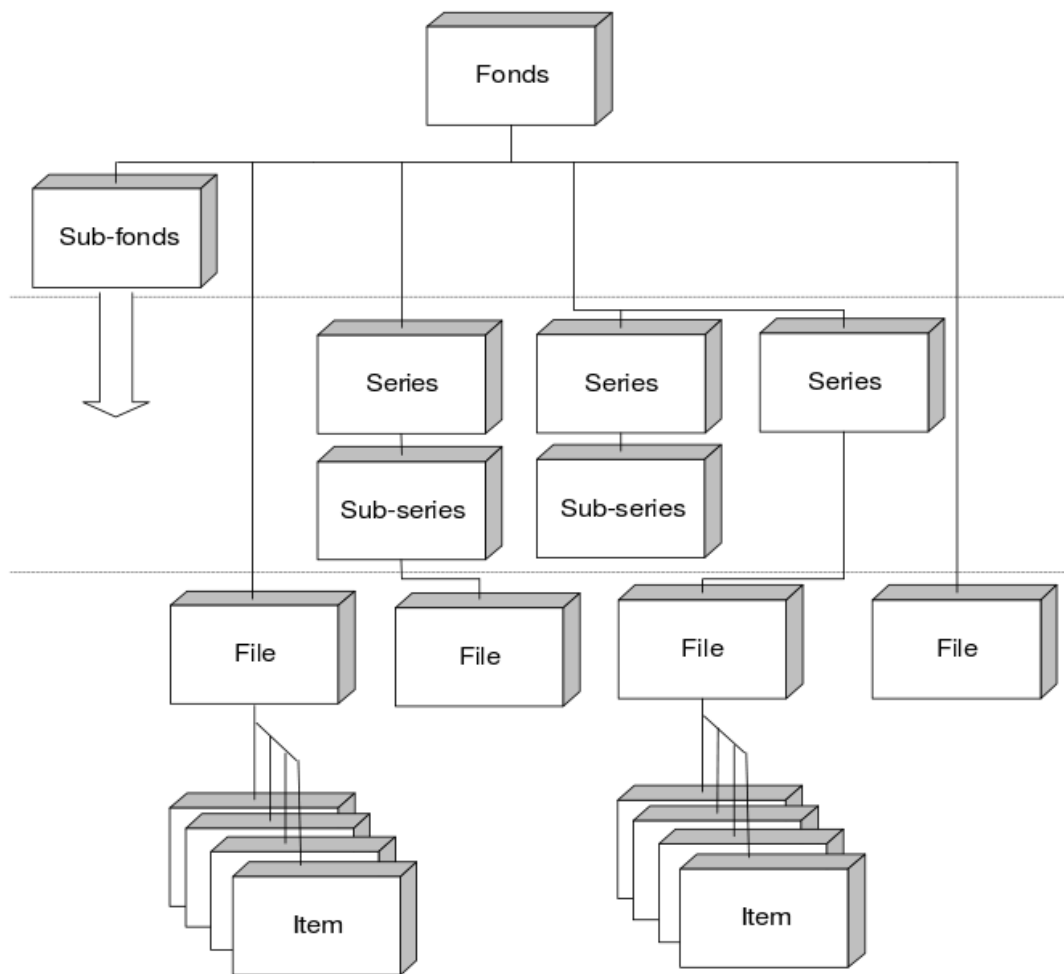


Figure 3 – Model of the levels of archival arrangement according to ISAD(G) (International Council of Archives (2000, 36)).

*Fonds* corresponds to “whole of the records” (International Council of Archives, 2000, 10), *series* are archival materials maintained as a unit because they result from the same accumulation or filing process, or the same activity; have a particular form; or because of some other relationship arising out of their creation, receipt, or use” (International Council of Archives, 2000, 11), *files* are “organized unit of documents grouped together either for current use by the creator or in the process of archival arrangement, because they relate to the same subject, activity, or transaction” (International Council of Archives, 2000, 10), and *items* are the “smallest intellectually indivisible archival unit” (International Council of Archives, 2000, 11) such as a letter, report, or photograph.

Typically, the description does not proceed to the level of individual items, mostly because

of the sheer quantities of material to be described. Furthermore, the extent and exact content of the actual descriptive information depends on available resources as well as the perceived importance of the fonds, the expected user interest, and the frequency of consultation (Franz, 2004, 89-90).

Just as the structure of the hierarchical descriptions may differ, the particular information provided on each level also varies; for example, regarding the provenance of records, archival description addresses questions regarding the creator of the assembled records the characteristics of the corporate body that created the materials, or the function or life activity that produced them (Fox et al., 1998, 6, 13).

Furthermore, archival descriptions may in principle incorporate a broad range of information regarding the documented activities, events, and functions, involved actors, locations and time-spans, or topics addressed, as well as the types of materials contained (Fox et al., 1998, 15-16). Information about the physical extent and condition (Fox et al., 1998, 15) as well as various administrative data such as access or use restrictions (Fox et al., 1998, 17) complement typical archival descriptions.

**Archival Aids** Commonly, the main *means of access* to the archival materials in an archive is via *archival aids* conveying archival descriptions as previously discussed and supported by the expertise of archivists. Archival aids comprise a variety of different descriptive tools such as catalogue records, inventory/registers, correspondent indexes, calendars of correspondences, published repository guides, or file plans, also including external aids such as indexes, file plans, data dictionaries and other tools incorporated in the records and created by the originators of the archival materials (Fox et al., 1998, 18; Franz, 2004, 37-38).

This study considers the finding aid<sup>37</sup> as the “workhorse of archival practice” (Gilliland-Swetland, 2001, 200) and most frequented archival aid by users and archivists alike. Archival *finding aids* serve as guides to the fonds by providing information about “the provenance of the archival materials and the original order in which they were arranged” (Gilliland-Swetland, 2001, 203) in a multilevel description and adhering to the previously discussed principles of the respect des fonds (Fox, 2001, 63). Typically, printed versions of finding aids serve archival users in the reading room and archivists in the reference service as means to identify relevant archival materials.

Creating an archival description is “first and foremost a delicate exercise in preserving evidence and making evidential values apparent to archival users in time and over time – making the materials intellectually available for interpretation, yet not providing that interpretation” and, at the same time, means “providing intellectual access to the factual information that archival materials contain on persons, places, and subjects contained in those materials” (Gilliland-Swetland, 2001, 201-202). The finding aid assumes three primary roles as a “tool that meets the needs of the archival materials being described by authenticating and documenting” as fonds, as a

---

<sup>37</sup> Finding aids are often also referred to as registers or inventories.

“management tool for use by the archivists”, and as “an information discovery and retrieval tool for making the evidence and information (...) available and comprehensible by archivists and users alike” (Gilliland-Swetland, 2001, 202).

Core information for the description of records in archival aids includes a title for the record, often a *title* of the respective file or series, *dates (Laufzeit)* from oldest item to newest item in the record, the *extent* of the record in terms of how many meters of shelving the contents of the record occupy, a brief *content description*, and *indices* such as person, places, or topics (Franz, 2004, 35-36). The guidelines as to what should be encoded in a title according to ISAD(G) are a good example for implicit knowledge potentially available in archival description: “For supplied titles, at the higher level, include the name of the creator of the records. At lower levels one may include, for example, the name of the author of the document and a term indicating the form of the material comprising the unit of description and, where appropriate, a phrase reflecting function, activity, subject, location, or theme” (International Council of Archives, 2000, 14). These brief instructions suggest an abundance of rich information to be encoded in titles at various levels of the descriptive tree. This example demonstrates that potentially a wealth of background knowledge will be necessarily accumulated during documentation by the archivist but, if the archivist chooses to do so, might be encoded only implicitly and isolated as plain text.

Another piece of core information includes *call numbers*, or any other kind of identifier, for a record, used to order the respective physical boxes with archival materials from the archive’s depot. Typically, searching for archival material with an archival finding aid means identifying the call numbers of records whose potential relevancy for one’s purpose is judged by the contextual descriptions.

The flexibility of the finding aid as a tool for information discovery and retrieval has increased through digital representations in online archival information systems regarding searching and displaying without giving up on the traditional hierarchical access and description (Gilliland-Swetland, 2001, 200, 203-205). However, as already discussed in the chapter “Related Research” (II), since archivists have been traditionally material-centric in their descriptive practices rather than focused on user needs in the design of the finding aids, their digital equivalents perpetuate these deficiencies (Gilliland-Swetland, 2001, 207).

As Gilliland-Swetland (2001, 208) has pointed out, finding aids are generally organized according to provenance and original order of the records but lack alternative access for secondary use. They have been further designed for use in an environment where the archivist not only mediates between user and archival material but also between user and the finding aid. Archivists explain the finding aid, help to identify relevant finding aids by drawing upon “their own contextual knowledge and reformulating users’ subject queries into provenance-based queries that can then be mapped onto the ways in which their archival holdings are arranged and described” (Gilliland-Swetland, 2001, 208). Finding aids outside the archive, for example on the Web, without the help of an archivist, can be confusing (Gilliland-Swetland, 2001, 208).

**Searching the Archive** This study focuses on state archives, and these adhere to the aforementioned archival principles. Since archival practices may vary, as previously stated, there is no fixed catalogue of inquiries. Nevertheless, archival principles entail several fundamental questions a researcher may ask when trying to identify, find or uncover relevant sources as evidence or information in an archive. There are certain questions and attributes which help to identify documents within the archive and which can be considered generally applicable. For example, according to Brenner-Wilczek et al. (2006, 65-92), important questions include: Where did a person work or of which organization or institution was he or she a member? How did the institution or group evolve (company, dynasties)? How were they organized? Which public or administrative body might have taken notice of or an interest in the person and his or her activities? Further inquiries may pertain to biographical information on a certain individual or indications of important phases in a person's life.

Such inquiries lead to administrative files of public bodies as well as other types of corporate bodies such as political parties or educational institutions. Personal estates, in particular, are important records in archives and contain a broad range of historical sources which are typically not found in other kinds of records such as diaries or correspondence (Brenner-Wilczek et al., 2006, 65-92).

Since the objective is to create an ontological model representing relevant user needs with regard to archives, such fundamental questions need to be considered during the interpretative process and incorporated while modelling research on the subject matters of interests. They are important "follow-up" questions which substantially support the identification of the concrete archival record or item within the archival structure. They provide information that is not directly or explicitly expressed in an inquiry but will help to relate the interest of the question to the information provided by the archival arrangement and description.

Certain pivotal entities and relationships between these entities can be derived from these questions together with the archival principles. For example, actors can be of great significance when seeking or identifying documents, and should therefore be incorporated into the general pattern as key elements. Similarly, responsibilities of actors for particular business activities appear to be important for the identification of documents in archives. The importance of such entities is not necessarily apparent from the inquiries since the questions typically specify what is needed but not how to obtain it.

This study will nevertheless as far as possible limit its scope to the requirements of interests formulated in inquiries. Additional context which does not directly pertain to the subject matter of the interest of the inquiry will be included only if this context is deemed essential in order to provide an adequate response in the context of general patterns or to enable important secondary or auxiliary questions. Context that may appear helpful but is not sufficiently backed by empirical evidence in the inquiries will not be included.

### 2.3 Summary

The interpretative analysis of the inquiries is guided by an epistemological framework comprised of basic and essential constituents of the historical and archival domain as well as general knowledge on the part of the interpreter. This epistemological framework underpins the interpretative analysis of the inquiries and allows for educated assumptions – where necessary – regarding the interest of the inquiries and its subject matter, as well as potentially relevant answers for the user.

In the scope of this study, history can be understood as an essentially empirical science which relies on the surviving traces and traditions of past human activity. Even though the epistemological framework delineated for the historical domain remains necessarily generic and broad, the essential research interest of historical inquiry ultimately boils down to the description, understanding, and explanation of past activities and events either actively conducted or passively experienced by actors. These actors may be individuals or any other kinds of social groups such as families, organizations or institutions. Some of the most basic and generic interests of historical inquiry then lie in questions about the motivation of actors for their actions as well as their plans in the widest sense such as laws, battle plans, orders or agendas, and whether and to what extent these plans have been executed and what their effects were.

Historical sources as found in archives constitute an important empirical reservoir for collecting factual and evidential information regarding these activities. The scientific form of historical inquiry deals with historical sources based on a tripartite historical method of heuristics, source critique and synthesis. Historical inquiry finds a literal form of expression in the user inquiries posed of archives. Formally, most inquiries can be expected to pertain to the heuristic stage of historical research.

Regarding epistemological issues in the interpretation itself in the light of historical science and theory of history, the approach taken here is conceived to be *meta-theoretical*, similar to Gardin (2003) in the domain of archaeology. While the approach does not consider specific types of historical science, it reflects basic patterns that appear applicable to general historical inquiry.

Since the inquiries are posed of archives, the archival domain is the second epistemological framework to determine the interpretative analysis and the ontological analysis of the inquiries. Historical primary sources which constitute the backbone of scientific historical inquiry are primarily found in archives. The selection of relevant adequate historical sources in terms of potential factual or evidential information is a key aspect of historical inquiry in archives. In principle, primary sources and the historical facts they contain – yet subject to questions about their truth, authenticity, or accuracy – serve as primary evidence for the historical knowledge created. Secondary sources such as scientific historical literature are typically not found in archives and therefore cannot be reasonably expected amongst the search results.<sup>38</sup> Archival

---

<sup>38</sup> Of course, if the inquiries exhibit a significant number of questions *explicitly* targeted at secondary literature then this demand would have to be reflected in the ontological model representing user needs.

principles such as *respect des fonds* and other tools such as finding aids further determine the possible interests of historical inquiries placed towards archives and how they could be addressed.

However, this study specifically approaches its topic not from a theoretical standpoint and does not seek to transpose the epistemology of the archival domain or historical domain into an ontology, but rather explicitly takes an empirical foundation as its starting point; that is, the user inquiries are the main source from which the epistemological framework gains substance and evolves, leading to the ontological model of user interests and their subject matter. Considering the infinite complexity of history, generalization and simplification in historical inquiry is paramount and intrinsic to the historical method and scientific endeavour.

The ontological model which formalizes and abstracts the subject matter of the interests of historical inquiry within archives serves to facilitate the selection of potentially relevant and useful sources. The query patterns then do not primarily answer inquiries directly by retrieving factual historical knowledge but their purpose in the current context is to identify historical sources which may contain information contributing to a response.

The last step towards this ontological model and of the interpretative analysis is the formalization and explication of the results of the interpretation. Both the interpretation and, at the same time, the translation of its results into an ontological structure are further determined by the design principles of the chosen target ontology, the CIDOC CRM, and the process of ontological modelling.

### 3 Ontological Modelling

The process of *ontological modelling* represents the last formal step in the ontological analysis of the inquiries, where the results of the interpretative analysis are formalized and explicated as general patterns (IV:1.2.2). The term *ontological modelling* has a bipartite meaning with regard to the methodological approach here: the term describes parts of the ontological analysis during the interpretative process, especially that of abstraction and formalization of the subject matter of the interests of an inquiry to common and generic patterns. Further, the term refers to the particular ontology of the CIDOC CRM, which has been evaluated and found adequate as the *means* for the ontological representation of the results of the interpretation. As an ontology the CIDOC CRM not only provides a set of concepts which can be used to represent entities but also entails, similar to the archival and historical domain, particular epistemological preconditions – such as its event-centric viewpoint – that influence the process and course of the interpretative analysis.

The following sections will provide an introduction to the CIDOC CRM and its way of thinking, thus elaborating further the idea of ontology. In particular, the adequacy of the epistemological preconditions of the CIDOC CRM for the historical domain will be assessed.

### 3.1 Modelling Methodology

The *International Committee for Documentation*<sup>39</sup> (CIDOC) of the *International Council of Museums*<sup>40</sup> (ICOM) has led the development of the *CIDOC Conceptual Reference Model*<sup>41</sup> (CRM) since 1996. In 2000, ICOM-CIDOC officially handed over the development of the CRM to the *CIDOC CRM Special Interest Group*<sup>42</sup> (SIG). In 2006 the CRM was accepted as the official standard ISO 21127.<sup>43</sup>

The *cultural heritage domain* exhibits a diverse spectrum of different sets of concepts representing different *points of view* on how to describe and conceptualize reality in the cultural and historical domain. They are embodied in various database schemas and documentation structures. As a *Conceptual Reference Model*, the role of the CRM is to provide a means for mutual comprehension and dialogue between domain experts based on a clearly defined formal and shared conceptualization of these seemingly different points of view (Doerr et al., 2003). In this regard, the CRM serves as an intellectual guide and common conceptual reference language to the creation and analysis of schemata, profiles or formats in a given cultural heritage domain. The intended purpose is to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information (Doerr and Iorizzo, 2008, 7). The *general scope* of the CRM can be described as all human activities and their products in the past as well as current evidence of such.

As a *formal ontology* (Sowa, 2000, 493), the CRM provides “a compact formulation of common concepts” (Doerr and Iorizzo, 2008, 20) representing and accommodating the underlying generic semantics of cultural heritage data and documentation structures. An ontology can be understood as formalized knowledge comprised of clearly defined concepts and relationships pertaining to possible states of affairs in a knowledge domain (Doerr, 2003).

The CRM was created bottom-up by re-engineering and integrating the common, dominant concepts identified in many different database schemata and documentation structures from many different disciplines and knowledge domains such as museums, archives and libraries (Doerr et al., 2007, 1). In other words, the CRM is not the product of inspiration but the result of an *empirical analysis* and observation of data structures and how experts use these structures for the purposes of argument. These observations are formalized as generic classes and relationships in a compact and nearly generic model. The CRM is therefore extensively based on *empirical evidence* regarding its adequacy, and it is fully multidisciplinary in terms of its development and contributors.

The CRM comprises *categorical knowledge* in the form of currently about 86 “classes” and 137 “relations”, terms which will be defined further below. These classes and relationships,

---

<sup>39</sup> <http://network.icom.museum/cidoc/>

<sup>40</sup> <http://icom.museum/>

<sup>41</sup> <http://cidoc-crm.org/>

<sup>42</sup> <http://network.icom.museum/cidoc/working-groups/crm-special-interest-group/>

<sup>43</sup> [http://www.iso.org/iso/catalogue\\_detail?csnumber=34424](http://www.iso.org/iso/catalogue_detail?csnumber=34424)

also called properties, are used as a “conceptual grid (...) superimposed to various possible states of affair” (Guarino and Pierdaniele, 1995) of the historical and cultural world. In other words, the CRM is a schema for factual knowledge which is strictly confined to a possible factual constitution of the past.

The following discussion of the modelling methodology of the CRM is based on Crofts et al. (2006, i-xviii) if not stated otherwise. References will be provided only for direct citations.

A *class* can be understood as a category for real-world items which share one or more common characteristics. The intended meaning of a class is called its *intension* and described in a *scope note* containing a textual description. The identity of classes in the CRM is represented by a label consisting of the letter 'E'<sup>44</sup> followed by a number and a name. These labels are mere mnemonics. The meaning and definition of CRM classes and properties is given solely in their scope notes. For example, *E21 Person* is a class and comprises, according to its scope note, “real persons who live or are assumed to have lived” (Crofts et al., 2006, 11). The items which are members of a class are called *instances* of that class and share the common characteristics described in the scope note. For example, you, the reader of this text, as well as the historical person John F. Kennedy but also legendary figures such as King Arthur who may have existed and are documented as possible historical figures would be instances of the class *E21 Person*.

A *property* describes a binary relationship between two classes. The binary relationship implies that a property has two distinct but related meanings in both directions. As in the case of classes, their identity is represented by a label beginning with 'P' for *property* followed by a number and a name. The name of a property is always given in both the active and passive voice since it is defined as inverse. For example, the categorical statement “*Person has current or former residence Place*” is equivalent to “*Place is current or former residence of Person*”. This categorical statement is typically expressed as one statement: “*E21 Person P74 has current or former residence (is current or former residence of): E53 Place*”.

Properties can be understood as verbs connecting a subject and an object. The subject then is the *domain* of a property and specifies the class for which the property is formally defined. The object is the *range* of the property and specifies all potential classes as values of that property. In the previous example, the domain of the property *P74 has current or former residence (is current or former residence of)* is the class *E21 Person* and the range is the class *E53 Place*. A class may be the domain or range of more than one property. As in the case of classes, the intended meaning of a property is called its *intension* and described in a *scope note* containing a textual description.

As an ontology the CRM represents categorical knowledge in the form of classes and properties while it allows for the aggregation of mostly *factual knowledge* about the past as instantiations of these classes and properties. Factual knowledge is comprised of *material facts*, which are propositions consisting of an instance of a property, also called a *factual relation*, which connects an instance of the class defined in the domain of the property and an instance of one of the classes defined in the range of the property (Degen et al., 2001). For example, the proposition “John F.

---

<sup>44</sup> The letter 'E' stands for *entity type*. The terms *entity type* and *class* will be used synonymously in the following.



Kennedy” (E21 Person) P74 has current or former residence (is current or former residence of): “The White House” (E53 Place) is a material fact.

The arguments in such a material fact can be understood as the subject, verb, and object, which are either *particulars* or *universals*. While a particular has no variations of itself, for example, *John F. Kennedy* or *The White House*, a universal has variations of itself, for example, *Person* or *Place*. In other words: A particular as an entity cannot have any instances while a universal as an entity can have instances. Both classes and properties are typically universals while instances of classes are typically particulars.

A special case of a material fact is that of *unitary relations*, which have only one argument. For example, the statements “John F. Kennedy exists” or “The White House exists” are material facts with a unitary relationship. These are *existential statements*.

Another important case of a material fact is the combination of a particular and a universal, called a *classification*, where the relationship is “instance of”. For example, the statements “John F. Kennedy is a Person” or “The White House is a Place” are classifications. These classifications make the particular “John F. Kennedy” an instance of the universal “E21 Person” (a class) and the particular “The White House” an instance of the universal “E53 Place” (a class).

The *isA* relationship is a principle of generalization between a concept and its broader concept (Brachman, 1983). The CRM employs this principle in order to organize the common concepts found in the cultural heritage domain into a compact hierarchy of broader and narrower classes and properties.

Two classes may be connected through an *isA* relationship, which means that one class is the *sub-class* of the other. A subclass specializes its *superclass* which means that all instances of the subclass are also instance of the superclass, the intension of the subclass is more restrictive, and the subclass inherits all properties from its superclass. The *isA* relationship is therefore transitive since characteristics are strictly inherited. For example, the classes E21 Person and E74 Group are both subclasses of E39 Actor. Furthermore, a class may be the subclass of more than one other, which is called *multiple inheritance*. For instance, the class E21 Person is also a subclass of E20 Biological Object, which means that the particular “John F. Kennedy” as an instance of E21 Person would also be an instance of E39 Actor and E20 Biological Object inheriting the semantics of these classes. Accordingly, two properties may also be connected through an *isA* relationship rendering one property the *sub-property* of the other. A sub-property specializes its *super-property*, which means that all instances of the sub-property are also instances of the super-property, the intension of the sub-property is more restrictive, and the sub-property inherits the domain and range of its super-property, including possible subclasses of these. This principle of an *isA* hierarchy allows for a reduction in the diversity and variety of concepts found in the cultural heritage domain to a relatively small and simple set of general and common concepts.

Furthermore, the CRM follows the principle of *minimality*, which means that the categorical knowledge defined in the CRM in the form of classes and properties are mostly *primitive* concepts.

These are concepts whose intension is declared without logical deduction from other concepts; for example, the concept “human” is primitive while “mother” is not if this concept is described as a “female human” with a “child” (Crofts et al., 2006, xii).

Ontologies such as the CRM that focus on generic semantics of data schemas and data structures are called *core ontologies* as opposed to *terminological ontologies*. The main difference between these two general types of ontologies is that terminological ontologies typically seek to capture large quantities of individual concepts in huge *isA* hierarchies in order to characterize or structure data in data fields or referred entities. Terminological ontologies define terms that typically appear as data in data fields. They structure the terms in order to describe their categorical properties and thus to facilitate definition and searching processes. To put it another way, following Doerr et al., terminological ontologies focus on the classification of terms by assigning them to concepts. Properties are mostly used to describe and qualify terms. For example, the terminology system *Unified Medical Language System*<sup>45</sup> (UMLS) comprises more than five million medical and pharmaceutical terms organized into a set of about one million concepts. On the other hand, the terminological system of UMLS is structured from about 130 generic types and 50 generic relationships defined in a core ontology (Doerr and Iorizzo, 2008, 5).

As a core ontology the CRM follows the insight that the understanding of cultural and historical contexts is primarily dependent on relationships and only secondarily on classification. For instance, whether a person is a politician or a criminal is less relevant than what this person may have done and whether or not these activities were of a criminal nature. And the latter question would be subject to the eye of the beholder. Classes and individuals only appear as qualities of relationships. Classes are declared only if they are required by the domain or range of a property, or if they are key concepts in a given domain. At the same time, it has been shown that only a small number of relationships are sufficient to express all the necessary semantics. On the one hand, this is due to the fact that simple classifications formally need only one relationship to connect an instance to a term from any rich terminological system. On the other hand, many intuitive relationships in the cultural and historical domain can be replaced by mediating events that, in turn, may be classified. For instance, “printed by” can be replaced with “produced by: Production (has type printing) carried out by”. As in the case of terms, relationships can be either specialized or generalized and thus reduced to small, meaningful sets. Doerr et al. (2007, 52) has pointed out that “the number of relationships in ontologies is orders of magnitudes smaller than that of classes and hence quite manageable” and that “a core ontology of ten to a hundred relationships can capture semantics of data structures across many domains”. Adhering to this principle, the research data is being analyzed primarily in terms of relationships. Concepts are only important when they motivate relationships.

The CRM follows the principle of *monotonicity* (Doerr, 2003, 82-83) which holds that any categorical statement in the CRM ontology or factual statement in a knowledge base following

---

<sup>45</sup> <http://www.nlm.nih.gov/research/umls/>

the CRM always has to remain valid and well-formed even when new categorical or factual statements are added to the CRM or to the knowledge base. This principle allows for the addition of new categorical or factual propositions which, however, need to be formally valid and well-formed, to a knowledge base following the CRM and thus facilitates the expression and representation of alternative opinions even if they appear to be scientifically contradictory; for example, by asserting two different fathers or mothers for one person.

Furthermore, the CRM adheres to the *Open World Assumption* which states that knowledge bases only hold incomplete knowledge regarding the domain of discourse described in the system. The assertion, therefore, that a fact may be missing does not necessarily indicate that it cannot exist in reality. In other words, one cannot inquire into items that apparently do “not have” some property. The *extension* of a class, defined as all potential real-life instances of that class adhering to the criteria of its intension, is therefore always understood as “open” and incomplete. Knowledge of the complete extension of a class is also an impossibility because the future may always bring new instances of that class. The instances of a class in a knowledge base are always only a subset of the extension of that class. In a cultural or historical context such an assumption is a fundamental necessity since the historical sources and records from the past are *always* regarded as incomplete.

To sum up the discussion of the essential modelling principles of the CRM, epistemologically, these principles – monotonicity, Open World Assumption and isA hierarchies – allow the CRM to manage two fundamental conditions of historical epistemology: “lack of knowledge” and “uncertainty”. In cases where details of a relationship between two entities or the particular nature of entities are unknown, the generalization can transfer uncertainty into less specific but *relative certainty*.

For example, based on the minutes of a meeting it might be uncertain whether a group actually created a specific document during that meeting. However, that they must have talked about the issue is generally certain since the minutes mention the document. One may assert with relative certainty that the group talked about the issue while, of course, someone else may assert that the group indeed created the document. The genuine task of the dedicated inquirer would then be to find and collect evidence on the specifics of this historical event and formulate a hypothesis.

The CRM is a disciplined way of aggregating knowledge but it is not a method via which to make decisions about possible truths. Furthermore, no *negative statements* are propagated by the CRM, which means that no statements are made about possible or probable states-of-affairs regarding the categorical or factual knowledge.

Now that this section has introduced fundamental concepts of the methodology of the CRM, the next section presents additional modelling concepts and the particular classes and properties the CRM provides for the purpose of modelling the domain of discourse touched upon by the inquiries.

### 3.2 Classes and Properties

The CIDOC CRM provides roughly 86 classes and 137 relationships along with scope notes describing their intended meaning and application. In this section, only selected classes and properties will be introduced in order to provide a general understanding of the workings of the CRM and its influence on the interpretative analysis of the inquiries. Entities incorporated in the general patterns will be introduced or further discussed in the chapter “Results” (V).

**Persistent and Temporary Things** The class *E1 CRM Entity* represents all things in the domain of discourse of the CRM (Crofts et al., 2006, 2). Nearly all 86 classes of the CRM are direct or indirect sub-classes of the class *E1 CRM Entity*.<sup>46</sup> The principal top-level entities of the CRM are shown in Figure 4. All other classes of the CRM can be understood as being specializations of one of these entities (Oldman, 2014, 8).

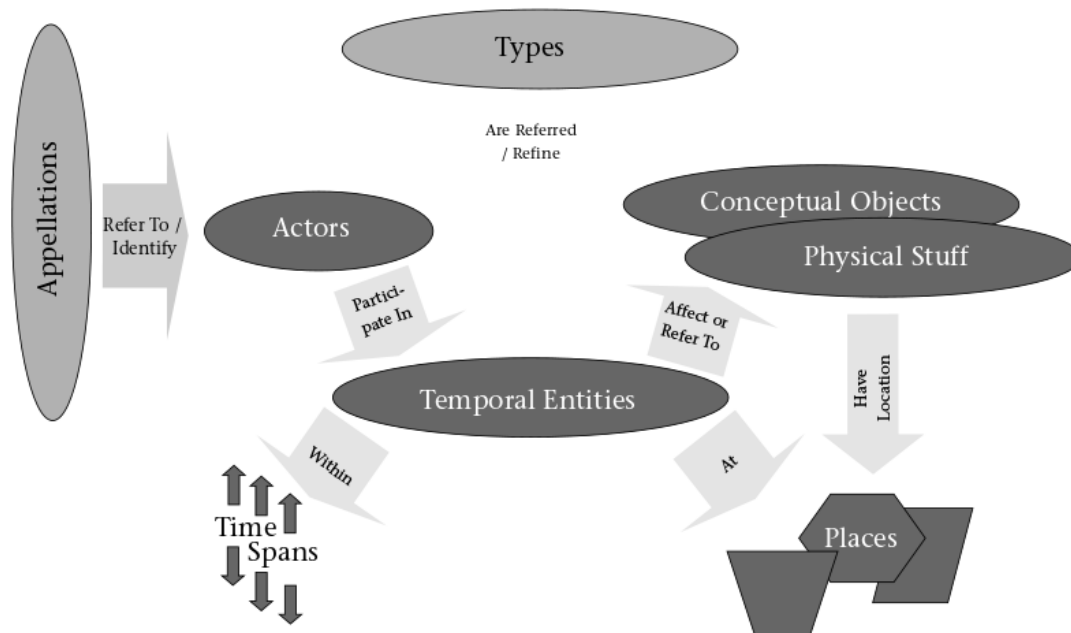


Figure 4 – The top-level classes of the CRM, taken from Doerr (2003, 85).

All these specializations are essentially split up into two principal trees of classes: persistent and temporal entities which primarily differ in terms of their behaviour in time. *Persistent entities*, represented by the class *E77 Persistent Item*, are things which endure over time (Crofts

<sup>46</sup> The only exception is the class *E59 Primitive Value*.

et al., 2006, 34). According to Masolo et al., in philosophy, they are also called *endurants* or *continuants*. Their identity persists over an indeterminate period of time even if their physical existence ends; for example, physical objects such as the Titanic or the World Trade Center, particular persons such as Albert Einstein or Ramesses II, or even ideas (Masolo et al., 2003, 10-11). This holds for anything which humans can remember. In other words, persistence means that their existence has the potential to span multiple, successive events with the same identity in which they can “participate”.

*Temporal entities*, on the other hand, represented by the class *E2 Temporal Entity*, are phenomena which occur only during a limited time frame (Crofts et al., 2006, 2). They exhibit temporal parts as time proceeds and their entirety is spread out over time. Their substance is the change, or even phases without change, implying change before and after. In language they correspond to the verbs or gerunds. According to Masolo et al., in philosophy they are also called *perdurants* or *occurrents*. Their existence is bound in time and ends after the fact, for example, natural events and activities of people (Masolo et al., 2003, 10-11). In other words, persistent items “can be repeatedly recognized within the duration of their existence by criteria” while temporal entities can only do so “by continuity or observation” (Crofts et al., 2006, 30).<sup>47</sup>

The interaction of instances from both trees of classes denominates a fundamental modelling decision of the CRM. Persistent entities such as people or things may either actively or passively participate in the occurrence of temporal entities; they may be affected by their occurrence, or witness it. Figure 4 also illustrates the pivotal role of temporal entities as a kind of hub connecting the others, such as actors, things, places, and time. This *event-based* approach is fundamental to the methodology of the CRM.

**Events and Activities (Temporal Entities)** One of the most important design principles of the CRM is to represent the past as a series of discrete events. Material and immaterial persistent entities are present at events either as concepts or via physical information carriers such as books or paintings. An object, more precisely objects made by man, only exists if some event has happened previously with this man-made object as its result or product. Natural things may exist as results of processes which cannot be cut into a discrete genesis; in other words, everything that has happened to an object – its administration, scholarly research, things that have happened to it in the past – as a part of events which have expressed part of a story. We can thus understand lives from the context of the relationships between things and where they might have encountered one another.

History, therefore, is conceptualized in the CRM as meetings of persistent entities through events in time and space (Doerr and Iorizzo, 2008, 7-8). As already discussed, the CRM also uses events in order to further reduce the number of relationships by grouping other entities around the notions of events. This view is empirically supported as many existing metadata

---

<sup>47</sup> The philosophical discussion of the distinction is confusing, even though it is intuitively very clear. The German philosophical terms *Seiendes* (persistent entities) and *Werdendes* (temporal entities) are more precise.

schemas indeed directly or indirectly support such event-based perspectives and representations of history (Doerr and Iorizzo, 2008, 8).<sup>48</sup>

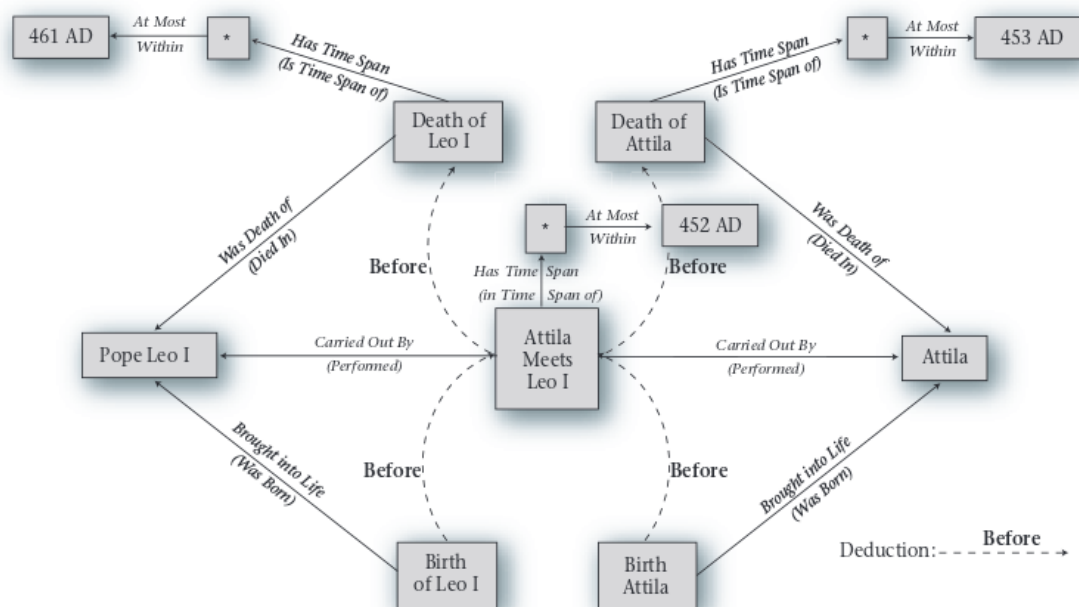


Figure 5 – The event-centric approach of the CRM, taken from Doerr (2003, 87).

Figure 5 shows an example of an event-centric description; that is, the meeting of Pope Leo I and Attila the Hun. The CRM indirectly connects the two particular persons, “Pope Leo I” and “Attila the Hun”, via the particular event with a mediating function, “Attila Meets Leo I”. The persons are said to have performed this event and must, therefore, have met each other. Even though there is no direct relationship between the two persons we can assert that they must have met each other at some place at some point in time.

The example also represents the birth and death events of both actors and thereby exemplifies the historical and temporal perspective of the CRM. As Lin et al. state, human thinking tends to compress complex paths of relationships into more simple ones while being aware of the whole context, a phenomenon also analyzed by Fauconnier and Turner (2003). In the CRM, these compressions are normalized by its event-based approach towards their implicit constituents. The concept of “mother” is an example where human thinking by-passes the occurrence of the “birth event”, while the CRM explicitly asserts the “birth event” in its wider context (Lin et al., 2008, 117).

<sup>48</sup> See, for example, mappings of the TEI to the CRM in Ore and Eide (2009).

Not only does the CRM unpack such compressions; the interpretative analysis of the inquiries identify implicit yet pivotal constituents – either in terms of classes or relationships – and renders them explicit by representing them in the ontological model. For this purpose, either existing entities of the CRM are evaluated and reused, or, where necessary, new ones are introduced by defining new classes or properties as organic extensions of the CRM. This process of formalizing and explicating implicit ontological entities during the interpretative analysis will be necessarily determined by the event-based approach, which therefore is one of the crucial epistemological influences derived from the CRM while analyzing the inquiries, as will be further discussed later on (IV:3.2).

The concept of a *period* (Doerr et al., 2010a), represented by the class *E4 Period* (Crofts et al., 2006, 3-4), is one of the most general temporal entities in the CRM. The two other relevant high-level temporal entities are *E5 Event*, and its sub-class *E7 Activity*. The class *E5 Event* comprises phenomena which result in recognizable “changes of states in cultural, social or physical systems” and have ramifications on *endurants* (IV:3.1), physical entities such as persons or things, or conceptual entities such as “ideas, concepts, products of the imagination or common names” (Crofts et al., 2006, 4). Instances of the class *E5 Event* are not purposefully caused by actors but by a *force majeure*, such as natural disasters, accidents, or celestial phenomena. In comparison to instances of *E7 Activity*, instances of the class *E5 Event* can be interpreted as “unintentional events” or “non-activities” which means they happen with or without the intentional participation or immediate causation of humans. To be sure, the class *E5 Event* is a generalization of the class *E7 Activity* which means that events neither require an activity nor exclude such. In other words, every activity is also an event but not every event is also an activity. The class *E7 Activity* then comprises any kinds of acts which have been carried out *intentionally* by individual people or groups (Crofts et al., 2006, 5-6). The meeting of Pope Leo I. and Attila the Hun is such an example of an instance of *E7 Activity*. Such deliberate and intentional actions are conceptualized as activities. It is not important whether the activity actually happened but only if there is a potential for deliberate and intentional action, and, of course, is deemed relevant knowledge from the point of view of documentation.

Through these temporal entities – and their numerous sub-classes, some of which will be introduced in the following paragraphs – persistent entities such as actors and things interact and are generally connected in time and space.

**Actors and Things (Persistent Entities)** Actors and things are both persistent entities and may be actively or passively present at events and activities during which they meet and interact.

The class *E39 Actor* comprises people either as individuals or as groups, who “have the potential to perform intentional actions for which they can be held responsible” (Crofts et al., 2006, 19). One of its two sub-classes is *E21 Person* which represents individual people who have lived or are assumed to have lived such as Karl Marx or Saint Paul (Crofts et al., 2006, 12). The other sub-class is *E74 Group* which represents groups of people, which means any gatherings

or organizations of more than two persons. Examples of groups include political parties such as the Communist Party of Germany (KPD), or the National Socialist German Workers' Party (NSDAP), or political offices such as the chancellor of the Federal Republic of Germany (FRG), or the party leader of the Greens, but also the concept of family (Crofts et al., 2006, 33).<sup>49</sup>

The ontological analysis of the inquiries focuses on deliberate and intentional actions (*bewusste Handlungen*) and the identification of these deliberately and intentionally acting entities. In this context, the class *E39 Actor* is also a good example of another important modelling principle applied during the analysis: when there is doubt as to whether a person or group has acted, then simply "an" actor is assumed. As already mentioned above, this allows us to deal with uncertainty in the model and the represented knowledge.

Actors then may be either actively or passively involved in events or activities. The high-level property *P11 had participant (participated in)* expresses this relationship (Crofts et al., 2006, 42). There are several sub-properties which specify the particular relationship between actor and event.

Other persistent entities such as physical things as well as conceptual objects such as ideas, concepts or products of the imagination may be present at or used in the context of an event or activity. For the purposes of this study, the two classes *E73 Information Object* and *E24 Physical Man-Made Thing* are the most important. Simply put, both classes and their respective sub-classes are in a relationship where a physical carrier, the physical man-made thing, such as a sheet of paper, carries some content (the information object), which in this example would be written on the sheet of paper. More formally, instances of the class *E73 Information Object* are identifiable but immaterial things in the sense of any kind of textual or visual contents whose existence and identity do not depend on the material properties of one particular carrier (Crofts et al., 2006, 32-33). The same text or image may be carved in stone and painted on a board. In other words, the class *E73 Information Object* comprises propositions in any symbolic form, such as words of a particular language, or letters and characters from an alphabet. Instances of *E24 Physical Man-Made Thing* are persistent and physical things which have been purposely produced by the activity of an actor (Crofts et al., 2006, 13). For example, a report written by a student is a physical man-made thing, while the contents of that report are an information object. This principle division allows for a description of contents independently from their physical carriers; for example, a text may be carried by the original sheet of paper on which it was written, while copies of the same text may exist, for example, as a microfiche or in digital form.

The presence of physical things and conceptual objects during events is expressed by the

---

<sup>49</sup> Institutions or groups of people with legal recognition as a group are represented by the class *E40 Legal Body* which is a sub-class of *E74 Group* (Crofts et al., 2006, 19). Legal bodies may act collectively and be held responsible like an individual agent. This class, however, is not used in the AKM: Differentiating between *E74 Group* and *E40 Legal Body* is difficult, not least because the CRM does not specify any distinct properties for *E40 Legal Body*. For the ontological representation of the subject matter of the interest of the inquiries, the distinction as to whether a group or legal body has or has not acted has not been relevant.



high-level property *P12 occurred in the presence of (was present at)* (Crofts et al., 2006, 42-43). The property also has several sub-properties specifying different notions of presence.

Apart from properties describing the relationship between persistent and temporal entities, many others exist such as part-whole relationships expressing for example the membership of actors in groups, or properties expressing a sequential succession of, for example, events.

Another fundamental example – shown in Figure 5 – is that of how persistent entities such as actors, physical things, and conceptual objects come into and leave, existence itself.

For actors, for example, there are the specific event classes *E67 Birth* (Crofts et al., 2006, 30) and *E69 Death* (Crofts et al., 2006, 31), during which persons come into existence, are born, and leave existence; that is, die a natural or unnatural death. Similar events pertain to the creation and destruction of physical things and conceptual objects.

The top-level property *P92 brought into existence (was brought into existence by)* (Crofts et al., 2006, 65) connects actors and things with the events via which they came into existence – mediated by the event sub-class *E63 Beginning of Existence* (Crofts et al., 2006, 29) – while the top-level property *P93 took out of existence (was taken out of existence by)* (Crofts et al., 2006, 65-66) connects actors and things with events of termination – mediated by the event sub-class *E64 End of Existence* (Crofts et al., 2006, 29). Again, these two top-level properties have several sub-properties which specify different qualities of the general relationship.

**Time and Place** The interaction of persistent and temporal entities occurs in time and space.

The CRM provides the necessary means to adequately represent the approximate time-span during which an event or activity occurred. The details of how time is represented in the CRM would exceed the scope of this study. An introduction to some of the difficulties of time representation in (an archaeological) context of the CRM can be found in Doerr et al. (2010b). The AKM refers to the available documentation and modelling options provided by the CRM for describing and specifying time-spans of temporal entities.

The class *E52 Time-Span* “comprises abstract temporal extents, in the sense of Galilean physics, having a beginning, an end and a duration” and is used “to define the temporal extent of instances of *E4 Period*, *E5 Event* and any other phenomena valid for a certain time” (Crofts et al., 2006, 23-24).

Instances of time-spans are always only approximations of the actual time-spans of events because historical knowledge is necessarily imperfect. The CRM, furthermore, provides properties to specify the beginning, end, and duration of a time-span, or to have a time-span fall within the duration of another. Even if a particular time-span is a vague approximation, as soon as this time-span is used for two events, they would be known to have occurred simultaneously (Crofts et al., 2006, 23-24).

Similar to time, the representation of geographical and spatial information also constitutes a complex topic in its own right. Here too, the details of modelling such information would exceed the scope of this study. Doerr and Hiebel (2013) provide a detailed discussion of related

issues. The AKM refers to the available documentation and modelling options provided by the CRM for describing and specifying geographical and spatial entities and information.

The class *E53 Place* “comprises extents in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter” (Crofts et al., 2006, 24-25). Examples of instances of *E53 Place* are typically immobile objects such as “buildings, cities, mountains, rivers, or dedicated geodetic marks”, but also mobile objects such as ships.

Finally, any CRM entity may be named, provided with an identifier, or given type information.

**Appellations and Types (Persistent Entities)** The CRM fundamentally distinguishes between instances representing real-world entities, particulars, and the names and identifiers which refer to them: “The relationship between entities and the identifiers that are used to refer to the entities, and the ambiguity of reference, are part of the historical reality that is to be described rather than to be resolved in advance” (Doerr and Iorizzo, 2008, 7).

The class *E41 Appellation* comprises “all proper names, words, phrases or codes, either meaningful or not, that are used or can be used to identify a specific instance of some class within a certain context” (Crofts et al., 2006, 20). Examples include names of actors, object identifiers, or place appellations such as addresses, or titles of documents.

In several cases, the range of a property is the class *E55 Type* which then can “refer in general to things of a certain kind” (Doerr, 2003, 85) and may further characterize or classify instances of other classes. The class *E55 Type* comprises “arbitrary concepts (universals) and provides a mechanism for organising them into a hierarchy” (Crofts et al., 2006, 25).

Most importantly, the “class *E55 Type* can be regarded as a metaclass (a class whose instances are universals), used to denote a user-defined specialization of some class or property of the Model, without introducing any additional formal properties for this specialization” (Crofts et al., 2006, 25). In the context of this study, the facility of *E55 Type* to express that an entity is of a kind; that is, that it has a type, will play an important role.

**Advantages of the CRM** The CRM is a core ontology for the cultural and historical domain for which other comparable ontologies and conceptual models exist, such as the *Suggested Upper Merged Ontology*<sup>50</sup> (SUMO) or the *Descriptive Ontology for Linguistic and Cognitive Engineering*<sup>51</sup> (DOLCE). The CRM has been chosen over such ontologies for a number of reasons, most of which have been already discussed in this section.

Most importantly, the CRM is firmly grounded in an “empirical analysis of real practice and local knowledge” (Oldman, 2014, 4) and therefore already provides empirical evidence regarding essential ontological entities relevant to the historical and cultural domain. Other ontologies lack such an empirical basis such as DOLCE, which is mostly based on intuition and provides a theoretically motivated logical description interpreting WordNet, an all-purpose

---

<sup>50</sup> <http://www.adampease.org/OP/>

<sup>51</sup> <http://www.1oa.istc.cnr.it/old/DOLCE.html>

dictionary free of context.

In other words, the CRM does not prescribe what to document but “brings into a homogeneous, integrated form what has already been documented” (Doerr et al., 2011). Since this study stresses the importance of empirically based representations of user needs, we will consider and evaluate the CRM for further use.

This empirical foundation also implies that the CRM is complementary to many metadata standards. For example, the CRM has been mapped to metadata standards such as the EAD, the most prominent encoding standard for archival aids. In the context of this study, these mappings allow for easy integration and comparison between the two ontologies. Furthermore, the CRM is aligned with other cultural heritage domain ontologies such as DOLCE or *Functional Requirements for Bibliographic Records* (FRBR) (Madison et al., 2009) and therefore has a broad semantic interoperability. Today, the CRM captures nearly generic concepts beyond its originally limited scope, and easily converges with models like ABC Harmony (Lagoze and Hunter, 2001), which was harmonized with the CRM, enriching it, and FRBR and the *Functional Requirements for Authority Data* (FRAD) (Bourdon et al., 2013), both now specializations of the CRM (Doerr and Iorizzo, 2008, 7).

Furthermore, since the scope of the CRM is the cultural heritage domain with a specific cultural and historical scope, its conceptualizations can be expected to be very close to the essential historical conceptualizations identified in the inquiries. For example, the event-centricity aligns well with the delineated historical ontology, and the CRM deals well with uncertainty intrinsic to any historical inquiry. Furthermore, the extensibility of the CRM has already been successfully demonstrated, such as in the case of CRMsci (Doerr et al., 2014), or CRMgeo (Doerr and Hiebel, 2013).

The CRM is relatively small in size and provides strong functional specifications which SUMO, for example, lacks. Finally, as already mentioned, the CRM is a ISO standard which provides the AKM with a strong and stable conceptual foundation.

### 3.3 Notation

The results of the ontological modelling process, the ontology AKM, will be delivered as a conceptual model and not as a data model, in text form as explanatory prose and with formal scope notes, and in graphical form as a range of diagrams. The notation used for representing the three different types of patterns in diagrams will be introduced. While the three principal types of patterns use the same basic notation, they also include, however, a few specific notational elements:

1. Diagrams for *general patterns* depict the principal classes and relationships which constitute a general pattern.
2. Diagrams for (categorical) *query patterns* summarize possible exemplary queries using given and wanted entities from a specific inquiry to instantiate elements from two general

patterns. *Query patterns* are meant to indicate possible yet not exhaustive queries, based on a particular user inquiry, for retrieving relevant documents or facts in order to answer the interest of a question.

3. Diagrams for *archival patterns* exemplify a complete instantiation of a general pattern without variables and based on empirical archival data.

The notational elements used in the diagrams are valid for all three types of diagram. The notation elements mainly used for the *general patterns* will now be introduced first; then the additional notational elements for the query and archival patterns will be discussed.

### Notation for General Patterns

Boxes in light-blue represent classes. The class name is given in the box. Solid simple arrows, *property arrows*, represent properties. The name of the property is given as the label of the arrow. The starting box of the property arrow indicates the intended *domain* and the end box the *range* of the property. In the example shown in Figure 6, the property *P2 has type (is type of)* points from the class *E1 CRM Entity* (its domain) to the class *E55 Type* (its range). In some cases, property arrows may represent more than one property, in which case property names are separated by “OR”, subsuming two or more separate property arrows. Property arrows pointing back to the same box indicate that the property may be asserted between two different instances of the same class. For example, the property *P148 has component (is component of)* may be asserted between two *E73 Information Objects*.

Sub-class relationships are represented by double arrows, *sub-class arrows*, between boxes. Solid double arrows represent a direct sub-class relationship while dashed double classes represent a sub-class relationship where at least one intermediate class has been omitted from the depiction. For example, in Figure 6, the class *E31 Document* is an immediate sub-class of the class *E73 Information Object*. The classes *E39 Actor* and *E73 Information Object*, on the other hand, are indirect sub-classes of the class *E1 CRM Entity*. Intermediate classes are typically omitted when they are not relevant to an explanation or to the study at hand.

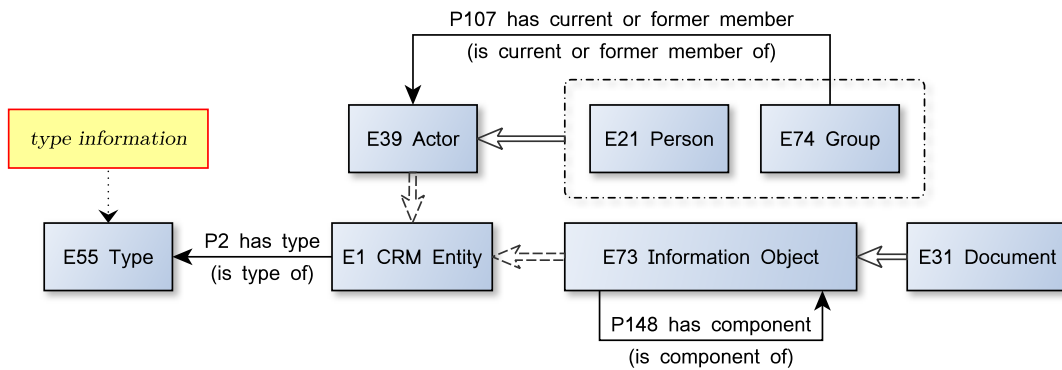


Figure 6 – Basic notation for general patterns.

In order to reduce the number of property arrows and sub-class arrows in the diagrams and thus increase readability, rectangles have been introduced as a shorthand. Rectangles may group two or more classes together in order to reduce the number of arrows it is necessary to display. Property arrows and sub-class arrows originating from or ending at the rectangle would pertain to all classes grouped. For example, in Figure 6, the sub-class property pertains to both classes *E21 Person* and *E74 Group*. If a property arrow or sub-class arrow is meant to pertain to only one specific class within the rectangle then the arrow will directly originate from or end at the appropriate box. In the example, the property *P107 has current or former member (is current or former member of)* only pertains to the class *E74 Group* since a person cannot *have* a member but only *be* a member of a group.

We must bear in mind that properties are inherited and therefore only usually displayed at the highest relevant level in the class hierarchy. For example, Figure 6, the property *P2 has type (is type of)* pertains not only to the class *E1 CRM Entity* but also to all other classes shown in the diagram since they are all sub-classes. Similarly, the property *P148 has component (is component of)* also pertains to the class *E31 Document*, and the property *P107 has current or former member (is current or former member of)* may be asserted not only between a group (*E74*) and an actor (*E39*) but also between a group (*E74*) and another group or person (*E21*) since both classes are sub-classes of *E39 Actor*.

In some cases, free text comments in yellow boxes with a red frame provide additional information on the use of specific classes or properties.

### Notation for Query Patterns and Archival Patterns

While the notation for general patterns suffices for representing the schema level of the AKM, the categorical query patterns and archival patterns also include the instance level and therefore introduce additional notational elements.

Diagrams showing query patterns or archival patterns may re-use the notational elements from diagrams for general patterns, such as comment fields or rectangles to group elements. The main difference between these and the general patterns is the incorporation of the *instance level*.

For this reason, boxes have two fields where the upper half contains the class name, the *class field*, and the lower half, the *instance field*, contains an instance of that class. Instances are always given in double quotes. For example, the class *E21 Person* shown in Figure 7 has a particular person as its instance, who has been rendered anonymous in this case by replacing the name of the person with “[person name]”.<sup>52</sup> The class *E73 Information Object* has the particular instance “Diary 1963 (contents)”.

The values given as the name of a particular in the example “[person name]” and “Diary from 1963 (contents)” only symbolically represent some unique identifier or method for approximating the identity of the particular. The value does not represent a specific method like string-matching to identify or find the named particular. In the context of query patterns, the names for particulars are typically taken from the original inquiry.

As another shorthand, type information as an instance of the class *E55 Type* may be given in the class field in square brackets below the class name. In the example in Figure 7, the box with the class *E73 Information Object* also contains the type information “Diary”. This is shorthand for *E73 Information Object – P2 has type (is type of): E55 Type “Diary”*.

The two classes *E12 Production* and *E65 Creation* have *variable* names in their instance fields, here “var-1” and “var-2”. Variables are always provided in italics starting with “var-” and followed by a running number. In cases where particular contexts are considered important in order to explain hypothetical queries but no particular instance data is available from the current example inquiry, a variable acts as a kind of place-holder which may be substituted by any particular instance during a hypothetical query. In other words, variables stand for any knowledge base content that fits into such a pattern. To a certain extent, they mediate between the given and the wanted. This means that the query will return all wanteds which have the indicated paths from the givens through any fitting value at the place of the variables. If two variable names are identical then their instances are also identical.

---

<sup>52</sup> Entities rendered anonymous are always given in square brackets.

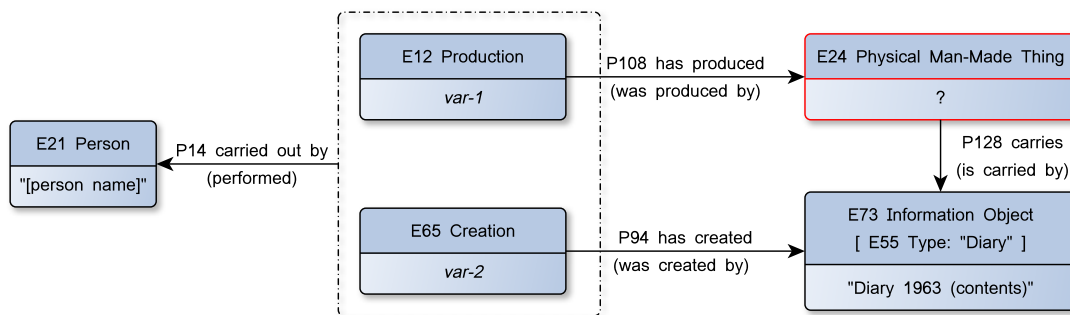


Figure 7 – Additional notational elements for query and archival patterns.

The main query target of a hypothetical query serving the current example inquiry is indicated by a “?” in the instance field and a red frame around the box. This syntax denotes that one or more instances of this class within the context of the particular query pattern can be considered an adequate and relevant result serving the main interest of the current example inquiry. Archival patterns do not have query targets since they are only exemplary representations of instance data from archival metadata.

The notation used in diagrams for query patterns is not meant to be explicit or unambiguous. As already discussed, query patterns are meant to exemplify possible queries adequate to theoretically satisfying the interest of an inquiry and typically subsume more than one possible query. This means, for example, that as an effect of inheritance (IV:3.1), potential queries may also utilize attributes of super-classes or super-properties which may not be explicitly represented in the query pattern. Furthermore, a general class or property entails that the potential sub-classes and sub-properties may also be used in a query. Finally, each box must be understood as having one or more instances; no cardinality is assumed.

The main difference between query patterns and *archival patterns* is that the former use the given entities from a specific example inquiry to instantiate a pattern while the latter completely instantiates general patterns with exemplary fictitious or real data. In other words, archival patterns do not contain variables.

Archival patterns have the purpose either to show how an inquiry could be met or which kinds of information would be retrievable, or to demonstrate how real life data would fit general patterns.

### 3.4 Summary

The CRM provides the means by which the results of the ontological analysis of the inquiries may be ontologically represented as general patterns.

This translative process from the particular level of the inquiries to the universal level of the AKM is called ontological modelling: relevant given and wanted particulars including relation-

ships interpretatively identified in the inquiries are further generalized and then formalized and explicated as universals in general patterns.

The general patterns represent categorical knowledge about the common user interests towards archives and also the subject matter of these interests. The historical reality they describe constitutes an adequately abstracted common context for identifying potentially relevant facts and historical sources in archives. The classes and properties used to create the general patterns are provided by the CRM. The main advantage of reusing these concepts is their foundation on extensive empirical evidence regarding their relevance for the representation of historical reality.

In addition, the methodology of the CRM further contributes to the epistemological framework as well as the process of ontological modelling, in particular the event-based approach to the representation of the past, the focus on relationships, and its facility to deal with uncertainty of knowledge about the past.

The basic frame of the CRM has been delineated. Only those classes and properties that are relevant for a fundamental understanding have been introduced here. All classes and properties which are relevant to representing the subject matter of the interest of the user inquiries will once again be presented and discussed in the next chapter (V).

## 4 Summary: Knowledge Engineering

For a study of the information behaviour of archival users, and particularly of their information needs, an epistemological investigation is essential. This refers to what and how the user seeks to learn from the inquiry. While Duff and Johnson (2001) analyze the information behaviour by looking at the structure of questions, this study ontologically analyzes the subject matter of the interests of users as found in their inquiries and in the context of the structure of the historical and archival knowledge domain itself.

*Knowledge Engineering* (KE), according to Feigenbaum and McCorduck (1983), involves “integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise”. Prominent examples of information systems as the result of KE are, for instance, automatic question-answering applications. While such sophisticated information systems are not the immediate objective of this, the outlined methodological approach is very much an endeavour in Knowledge Engineering. The iterative analysis of inquiries comprising interpretation and subsequent ontological modelling extracts and formalizes knowledge structures underpinning the highly complex act of finding and discovering historical sources in archives. The ultimate goal of the ontological model, the *Archival Knowledge Model* (AKM), is indeed to facilitate and enhance such archival search and retrieval operations which typically demand very “high level of human expertise” as a result of this analysis.

The CRM is an ontology that allows for the description of historical facts in terms of relationships between universals. The ontology is the result of an empirical analysis of existing conceptualizations of the cultural and historical world in the form of metadata structures. As



such, this study also conducts an empirical analysis of existing conceptualizations by analyzing empirical data; that is, archival user inquiries. The AKM evaluates and re-uses the empirically sound concepts of the CRM for its own purpose. The subject matter of common interests identified is explicated and formalized by means of the CRM; that is, by evaluating its semantics and utilizing its classes and relationships where appropriate in order to formulate general patterns. Where the CRM is found inappropriate during the course of modelling inquiries, it is specialized by new, compatible classes and properties.

New classes are only introduced if they motivate a relevant and indispensable new relationship. Otherwise, in order to refine or further characterize instances of classes and which do not legitimize the introduction of a new class, the class *E55 Type* is to be used. New classes and properties will be motivated and described when first introduced in the discussion in the following chapter (V). New classes will be prefixed with “C<sub>akm</sub>” where “C” stands for class and new properties will be prefixed with “R<sub>akm</sub>” where “R” stands for relationship.

Furthermore, the CRM demonstrates that a “bottom-up” engineering approach can lead to a “nearly generic” information model with a great deal of potential, wide applicability, and expressive power as compared to “top-down” approaches (Doerr and Iorizzo, 2008, 8). This study embraces a similar approach insofar as a generic ontological model is constructed “bottom-up” starting from the broad empirical basis of the inquiries. The AKM does not derive its concepts from specific theory but utilizes the outlined epistemological framework only for additional guidance.

The AKM is an adequate representation of a past reality which represents the domain of discourse of the inquiries and the subject matter of their interest towards archives while, however, not recreating an “objective” past reality. The model is adequate to its purpose and thus necessarily “subjective”. In this regard, the AKM does not differ from the writings of historians and the documentation created by archives, which are all appropriate representations for the particular interest of their creators. The AKM and its general patterns represent purposefully created ontological descriptions of possible pasts, not of “real” ones, and a partial representation of particular practices in the historical domain (Krausz, 1991).

The AKM as an ontology can then be understood as an ontological representation of the “real questions” extracted from the inquiries. The general patterns reasonably reduce the complexity of the interest of inquiries and their domain of discourse to material facts, and define an adequate level of detail as a compromise between complexity and simplicity as well as between the needs of archive users and the potential information archives typically provide. Each general pattern describes a particular close-up view of the historical context; that is, of the subject matter of interests relevant to particular sets of user inquiries.

Figure 8 stresses the role of the AKM and its general patterns as a mediator between the historical domain; that is, the domain of discourse of the user inquiries, and the archival domain. Query patterns and archival patterns are instantiated general patterns. Both types of instantiated patterns solely serve demonstrative purposes in order to exemplify the relationship between

the ontology, the AKM, and the respective knowledge domain. The query patterns describe the interest of user inquiries while archival patterns explicate particular knowledge from archival aids which is relevant to user inquiries.

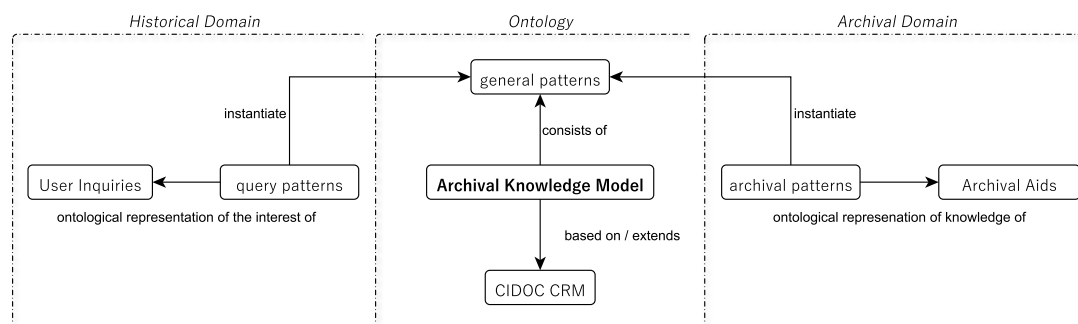


Figure 8 – The AKM as a mediator between the historical and the archival domain.

In this regard, the general patterns constitute formulas for potential queries, called query patterns, towards archival aids such as finding aids which themselves provide paths to material within the archive. The effects; that is, the particulars found in the inquiries which may instantiate general patterns, are limitless, while the types of potential historical sources are not. The question is not only whether particular query patterns lead to a direct answer to a question but whether the query patterns lead to source materials that potentially contain an answer to a question. Query patterns therefore describe the relationship between user inquiries and archival aids as partially instantiated general patterns. The ontology, the AKM, then describes a historical reality and constitutes an empirically founded and controlled means for discovering potentially relevant materials in archives. As such, the AKM is intended to supplement data structures for archival finding aids rather than to replace them. The relationship between archival aids and the AKM is described and exemplified by archival patterns.

One of the hypotheses of this study is that shared interests can be identified in inquiries posed of archives and that the subject matter of these interests can be generalized and formalized as described in this chapter. The interpretative analysis conducted here is necessarily a process of abstraction and simplification gradually stepping beyond the level of particulars as found in the user inquiries towards common universals forming general patterns. The single formal steps of the interpretative analysis are conducted iteratively. In particular, the transition from the implicit results of the interpretation to their explicit and formal representation is very much a process of *modelling* (McCarty, 2005).

The interpretative analysis begins with the extraction of single inquiries from the collected user and case files and the recording of single questions and associated contextual inform-

ation in a table. The individual questions are then categorized according to their primary interest (resource-discovery, fact-finding, or non-discovery questions). From here, only resource-discovery and fact-finding questions and their associated contextual information are analyzed further. The next step is the identification and extraction of the given and wanted entities also recorded in the table.

Based on these initial analytical steps, the shared interests of the inquiries and their subject matter are interpreted further. The process of ontological analysis integrates the given and wanted entities into a semantic network of relationships and further consolidates this network by generalizing and supplementing the ontological representation with essential entities from the archival and historical domain not apparent from the inquiries.

When interpreting the questions, the uncertainty factor in terms of adequate answers is addressed by abstracting to the most general but also least uncertain ontological entities. In other words, in the case of “fuzzy” or indeterminate information requests, querying adequate higher abstractions increases the recall of the query. This means that all possible hits are contained in the answer by an information system, albeit together with more, unrelated, information.

The complexity of the inherent phenomena in the inquiries are further reduced to an adequate level of abstraction by deconstructing the underlying past “reality” (Fauconnier and Turner, 2003) and then formalizing relevant entities using concepts of the CRM. The phenomena found in inquiries submitted to archives are potentially endless and must be selectively abstracted to the most relevant. The user inquiries constitute the primary empirical foundation for identifying common interests and primarily determine which ontological structures are relevant. The interpretative analysis, however, is further guided by an epistemological framework comprising knowledge from the archival and historical domain as well as common knowledge of the constitution and operation of our world. The CRM, furthermore, provides empirically grounded conceptualizations for representing historical realities.

Let us consider an example of how an epistemological framework guides the interpretation using a question inquiring as to the motivation of an actor for conducting some activity (note that the abstract formulation also constitutes an example of intermediate and often implicit steps of abstraction of the particulars found in the inquiries). One could assert a new property in order to express that a particular document contains the motivation of an actor in relation to an activity. However, while on the one hand this kind of information is not available in archival documentation and cannot reasonably be expected, on the other and even more importantly, this kind of knowledge regarding motivation must be extracted and understood by the historian. The answer to this research question is not direct. It is the follow-up question relevant in the current epistemological context that asks for particular kinds of documents pertaining to objective historical conditions and contexts. The objective consists in relevant documents which might contain an answer to the initial question. One important question during the analysis, therefore, asks what is observable and what can be reasonably expected from an archive. The motivation or other psychological conditions are not directly observable and they are not described and

recorded in archival information systems. Another key question, therefore, is how to answer such questions within the boundaries of the given archival context.

This also means that the focus is not primarily directed at whether particular material facts can answer a question directly, but rather on whether particular query patterns allow for the procurement of material, primary and secondary materials providing an answer to the question.

Another example is the event-centric approach of the CRM: When representing the context from which potentially relevant documents might have emerged, events and activities will play a pivotal role. When a question asks whether an actor has been at a particular place (a fact-finding question) then the AKM will not relate the actor directly to the place but will assert an event or activity during which the actor was present and which occurred at said place. Again, events have a pivotal role in the conceptualization of the historical past as relevant to the interest of the inquiries.

The guiding question particularly during the ontological modelling is whether potential query patterns allow for the identification and retrieval of historical sources likely to provide evidential or factual information relevant to the interest of an inquiry. In other words: Which are probable and adequate answers to the question while also considering the epistemology of the historical domain and the information potential of the archival domain?

The next chapter will present in detail the results of the practical application of the interpretative analysis.

## Chapter V

# Analysis and Results

The three main sections of this chapter present the results of the study. Beginning with the primary outcome of the practical application of the interpretative analysis (IV) to the user inquiries (III), this chapter will also elucidate the results of the *Archival Knowledge Model* (AKM) and its general patterns. Since the methodological approach itself is subject to the research interest of this study, its practical application and adaptation will be subsequently discussed in the first main section on the “Interpretative Analysis” (V:1). The second main section, “General Patterns” (V:2), presents and discusses the outcomes of the interpretative analysis and major results of the entire study.<sup>53</sup> The last section, “Summary: Archival Knowledge Model” (V:3), then summarizes the findings and the general patterns.

## 1 Interpretative Analysis

The interpretative approach to the analysis of user inquiries has been introduced in detail in the previous chapter (IV). Formally, the interpretative analysis is structured into an initial linguistic analysis followed by an extensive ontological analysis which includes ontological modelling. In practice, the various steps are conducted iteratively and cannot always be viewed in isolation from each other.

This section will discuss the results of significant steps of the interpretative analysis, starting with the categorization of the user inquiries according to types of questions (V:1.1) and the investigation of the given and wanted entities (V:1.2). This will subsequently lead on to the discussion of the interpretation process. Here, the further interpretative investigation of the interest of the inquiries (V:1.3) will be followed by a discussion of the ontological core framework (V:1.4) as the essential outcome of the interpretative process and transition to the ontological model. The ontological model (V:1.5, V:1.6) will then be outlined further, preparing the in-depth presentation of the general patterns constituting the primary product of the interpretative analysis. These are discussed separately in the subsequent section (V:2).

### 1.1 Types of Questions

The initial extraction of 762 single inquiries (IV:3), that is, of single questions and associated contextual information, from the collected archival reference questions is followed by the categorization of each single question according to the most recognizable information need expressed in the inquiry. Examples of such key information needs particularly include finding

---

<sup>53</sup> Early versions of the general patterns have been discussed in Hennieke (2014, 2013, 2011b,a).

archival material and identifying facts, or instigating other kinds of requests. The type of question assigned to a single inquiry indicates these principle information needs.

The three primary categories of question types (IV:1.1.1) are “resource-discovery”, “fact-finding”, and “non-discovery” questions. The category “resource-discovery” includes questions of the type: material-finding, specific type, specific item, research question, and consultation. The category “fact-finding” includes only questions that seek to identify factual information. The category “non-discovery”, encoded for the study despite not being subject here to further analysis of given and wanted entities, includes questions that are generally of an administrative or directional nature, service requests, or serve the purpose of improving the user’s knowledge in terms of technical, procedural, or methodological aspects regarding the usage of archives or research in general, here referred to as “user education” inquiries.

Figure 9 shows the total and relative distribution among the three primary categories of question types in the entire sample. Of all questions, 364 (48%) fall into the category “resource-discovery”. In combination with “fact-finding” inquiries, which occur in 112 cases (15%), substantial questions constitute 63% of all inquiries in the sample, rendering a total of 476. Non-discovery questions amount to 286 in the whole sample (38%).

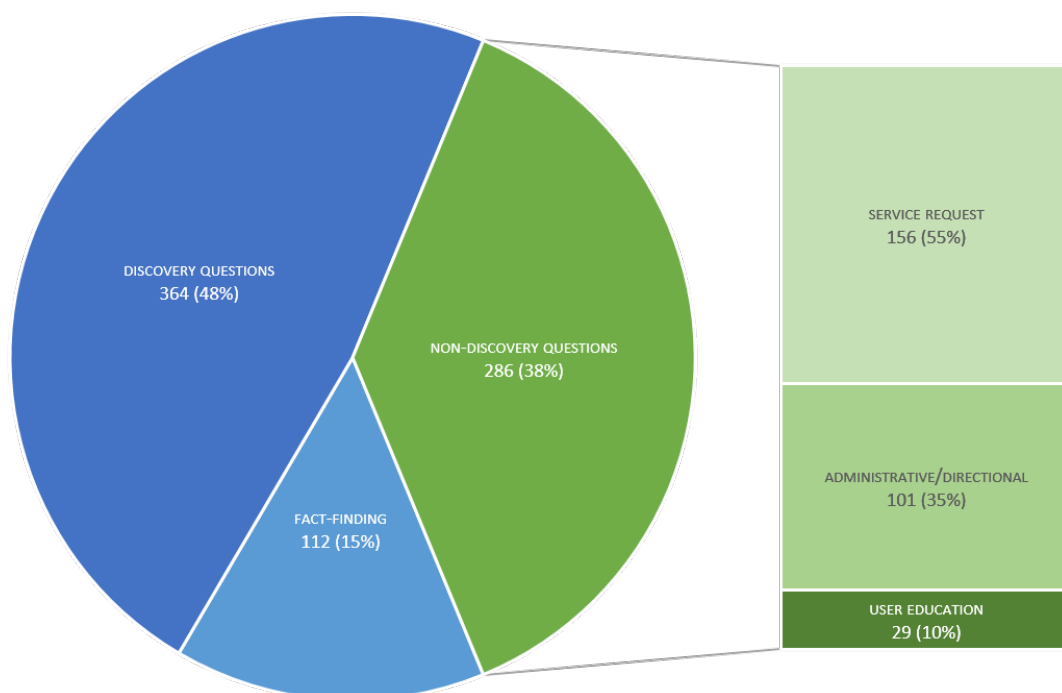


Figure 9 – Primary categories of question types detailing non-discovery questions.

Figure 9 also includes a listing of the three different types of “non-discovery” questions.<sup>54</sup> Within this category, service requests amount to 156 (55%) occurrences and constitute more than half of all non-discovery inquiries. Administrative / directional questions appear 101 times (35%) while only 29 questions (10%) were found to address aspects of “user education”.

Figure 10 lists the distribution of the five different kinds of “resource-discovery” questions.<sup>55</sup> Within this category, material-finding questions rank highest with 146 (40%) occurrences, followed by specific type and specific item questions, both of which amount to 70 cases each (19%). Research questions were identified in 47 cases (13%) and consultation questions in 31 (8.5%).

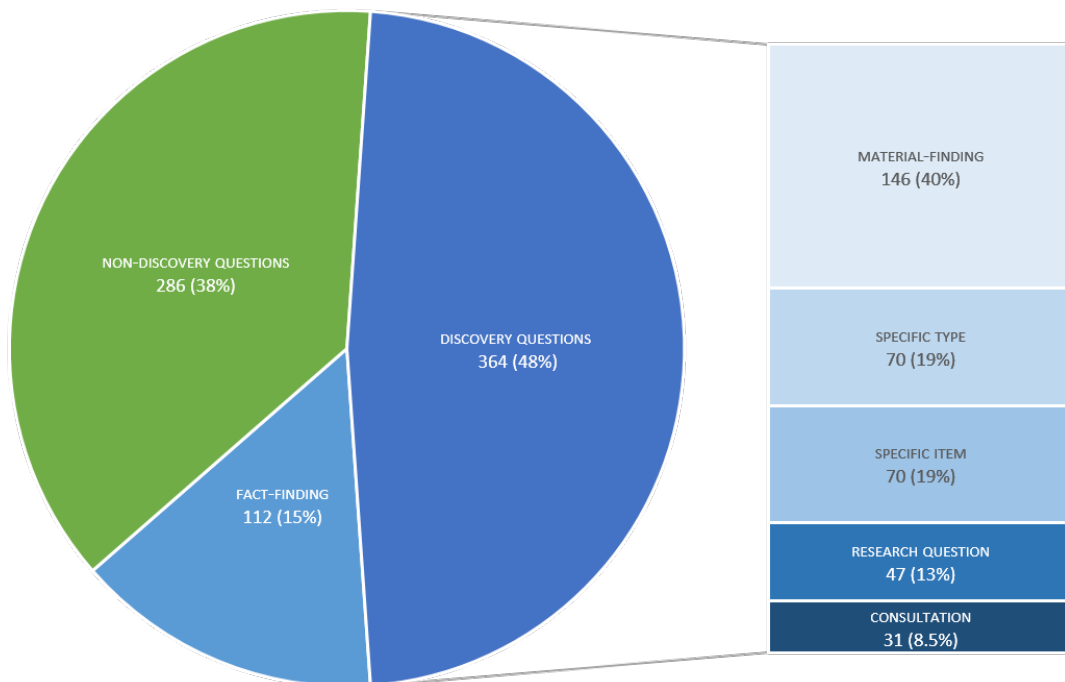


Figure 10 – Primary categories of question types detailing discovery questions.

Comparing these figures to the results of Duff and Johnson (2001) is difficult because the study at hand defines the scope of the question types differently to a certain degree (IV:1.1.1). The relative amount of fact-finding questions appears to be similar in both studies, however,

<sup>54</sup> The percentages provided for service request, administrative / directional, and user education are relative to the non-discovery category.

<sup>55</sup> The percentages provided for research question, material-finding, specific type, specific item, and consultation are relative to the discovery category.

with Duff and Johnson (2001) finding 10% and the study at hand 15%. Duff and Johnson (2001) counted only “material-finding”, “specific form”, and “known item” questions in the “resource-discovery” category, which amounted to 29% of all questions in their sample. Consultation questions, which constitute 10% in the sample of Duff and Johnson (2001) are excluded from the category “resource discovery”. Furthermore, some user education inquiries in the case of Duff and Johnson (2001) would in the context of the study at hand be more appropriate to the category “resource discovery”. Therefore, *cum grano salis*, the number of “resource-discovery” inquiries could be estimated at around 10% higher here than in the study of Duff and Johnson (2001).

Figure 11 considers only “substantial questions”, that is, “resource discovery” and “fact-finding” questions, and shows their distribution per sample.<sup>56</sup> Comparing the relative distribution of each question type between each sample, significant differences are unveiled in the cases of “material-finding” and “fact-finding” inquiries. The relative amount of fact-finding questions is higher in the sample from the National Archives of Norway than in the one from the German Federal Archives, while the relative amount of material-finding is higher in the latter. Since the relative amount of research questions is also higher in the BArch sample, one explanation might be that research questions often subsequently lead to further “material-finding” questions in the same inquiry.

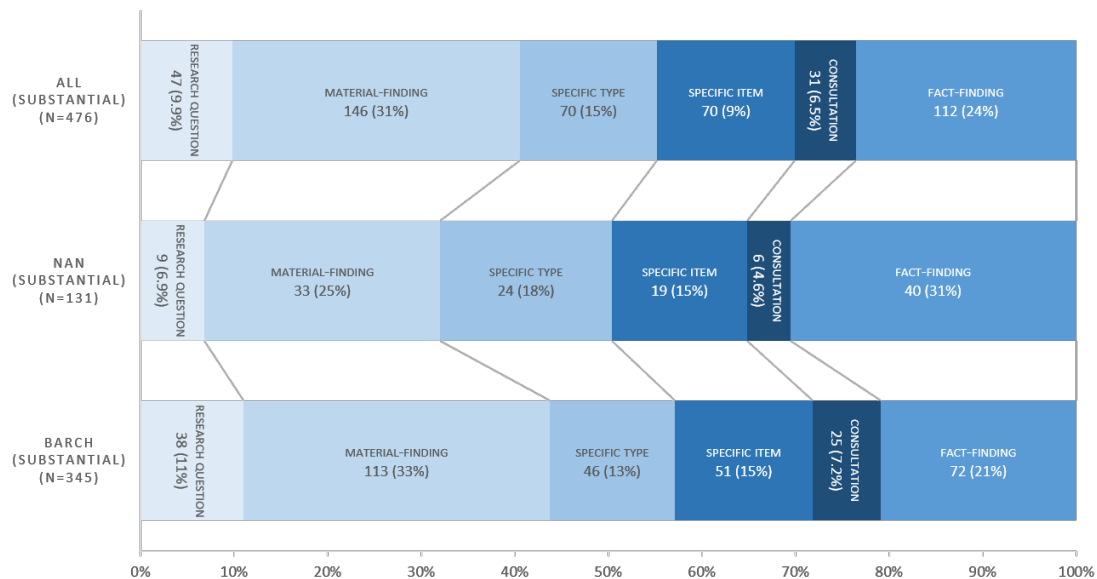


Figure 11 – Substantial types of questions per sample.

<sup>56</sup> Note that, in Figure 11, non-discovery inquiries are excluded from the numbers.



In total, the sample for this study comprises 762 single questions, of which 364 (48%) were *resource discovery* questions, 112 (15%) *fact-finding* questions, and 286 (38%) *non-discovery* questions. The subsequent analysis takes into account only inquiries of the type *discovery* and *fact-finding*. Together, these 476 single questions (63%) and their associated contextual information will be subject to further analysis, beginning with the investigation of the referenced entities; that is, the given and wanted entities.

## 1.2 Referenced Entities

The principal framework of *Givens* and *Wanted*s as found in Duff and Johnson (2001) and its extension to the idea of *given entities* and *wanted entities* were introduced in the previous chapter (IV:1.2.1). The following two sections present the analysis of the given and wanted entities identified in the single questions and associated contextual information.

### 1.2.1 Given Entities

The *given entities* represent those entities referenced by the user in the question and associated contextual information in order to further describe and qualify the interest of the question (wanted entity).

As previously mentioned, this study does not primarily seek to quantify how many entities of the same kind are referenced but rather how many different kinds of entities are provided in inquiries. For this reason, the analysis only counted one distinct type of a given entity per inquiry. This means, for example, that even if more than one particular place is mentioned in an inquiry, only one particular place is counted by the analysis for this specific question.

The given entities are grouped into seven main categories which grew iteratively from the analysis and comprise *Actors* (persons and groups), *Documents*, *Times* (periods and time-spans), *Places*, *Events* (unintentional events and activities), *Things*, and *Other Entities* (identifiers and general contexts). In each case, the analysis fundamentally distinguishes between particular entities and types or kinds of entities. This differentiation allows for an assessment as to the specificity of knowledge transposed in inquiries and, at the same time, paves the way for the further ontological modelling.

Table 1 – Given Entities

Actor	Description
<i>particular person</i>	A particular person <sup>57</sup> is typically referenced by a proper name but often also indirectly by a phrase. Examples are "Walther Funk", "Göring", "[the Minister of the Interior] Frick", or "my father".

<sup>57</sup> Corresponds to "name of person / corporate / entity" in Duff and Johnson (2001).

<i>type of person</i>	A type of person <sup>58</sup> refers to the occupation or profession, the social status, role or function, or also state of membership in a group. Examples are "[he is a] doctor", "ambassador", "photographer", "owner [of a printing office]", "philatelic official", or "member [of the KPD]".
<i>particular group</i>	A particular group is a gathering of more than one person able to collectively act or form a will or decision and official or administrative offices which commonly have only one representative at any given point in time. Examples are "the government of the German Reich", "the Reichstag", "the Federal President", "the leadership of the SED", "the SED", "the Göring family", or "the fraction of The Greens".
<i>type of group</i>	A type of group is, for example, "foreign citizens", "official authorities", "companies", "revolutionary movements", "German leftist intellectuals", "German authorities", or "communist parties".
<b>Document</b>	
<b>Description</b>	
<i>particular document</i>	A particular document loosely corresponds to "form" in Duff and Johnson (2001) and can be characterized by their form such as "the photograph [of your grand-mother]", "the map [of Berlin]", or "the file 20/03 on film roll", or by their function such as "the cadre files [of your grand-father]", "the participants lists [of the conference]", or "the report [of the trip to France]". Documents also include collections of various kinds such "the party archive [of the SED]", "the personal estate [of a particular actor]", or "the journal '[name]'".
<i>type of document</i>	A type of document then includes kinds of documents according to their nature such as "photographs", or "drawings" in general, or to their function "documents with personal assessments", "personal remnants", "collectors' magazines", "petitions (of the populace of the GDR)", or "delegates lists (of the party conferences of the KPD)". Examples of types of collections would be "files (of the KPD)", or "the Norwegian police archives".
<b>Time</b>	
<b>Description</b>	
<i>particular period</i>	A <i>particular period</i> is, for example, "National Socialism", "the Weimar Republic", "the Eighties", "the early years of the GDR", "early Eighties", or "Industrialization".
<i>type of period</i>	A <i>type of period</i> is, for example, "ruling system", "artistic style period", or "political period". These kinds of entities do <i>not appear</i> in the sample.

---

<sup>58</sup> Corresponds to "occupation or profession" in Duff and Johnson (2001).

*time-span* A *time-span* cannot be distinguished into types and particulars. Time-spans often characterize activities or events such as births or deaths. Examples are "1975-1985", "22. Nov. 1891", "September 1986", "1989", "before August 1944", "night of October 22/23 1943", or "Summer 1987".

**Place** **Description**

*particular place* A *particular place* is a current or historical specifically referenced geographical entity. Examples are "the GDR", "reunited Germany", "Berlin", "Siberia", "HH"<sup>59</sup>, or "Central Europe".

*type of place* A *type of place* is a general type of geographical entity such as "deployment locations", "villages", "inland", "coasts", "borders", or "industrial areas".

**Event** **Description**

*particular event* A *particular event* primarily includes natural "non-activities" such as "the aircraft accident [of 22 February 1988]", "the flood [of 1962 in Hamburg]", or the birth or natural death of a person if explicitly expressed in the text.

*type of event* A *type of event* includes types of natural "non-activities" such as "natural disasters", "accidents", "flooding", or "diseases". The implicit reference by dates to the birth or death of a person are counted as types of event. Finally, type of event subsumes "conditions" (*Zustände*) or "temporary states of a thing or actor" such as "to be a guest at [someone's house]", or "to be a resident of [a country]".

*particular activity* A *particular activity* is an activity intentionally carried out by an actor or in which an actor either actively or passively participated. Examples are "the denazification (of [person name])", "the laying down (of a wreath of flowers)", "the population census", "the transfer of custody (of the files)", "the trip [to France]", "the first meeting [of the AG]", or "the planning (for the 'Day of Potsdam')".

*type of activity* A *type of activity* is a general kind of activity such as "to raise official protest", "nationalizations", "denazification [in Germany]", "social policy [in the GDR]", "party conferences [of the KPD on the Reichsebene]", "meetings [of the Reichstag]", or "granting [of Norwegian citizenship]".

**Thing** **Description**

*particular thing* A *particular thing* is, for example, "the medal [awarded to my grandfather]".

*type of thing* A *type of thing* could be, for example, "[Norwegian] ships".

---

<sup>59</sup> Stands for "Hansestadt Hamburg", hanseatic city of Hamburg.

Other	Description
<i>identifier</i>	The category <i>identifier</i> subsumes all kinds of citations, call numbers, titles, or other kinds of labels. Identifiers are often given in combination with particular documents.
<i>general context</i>	The category <i>general context</i> captures all things and phenomena which have not been further analyzed but which play an active or passive role within a given sphere of interaction or cultural and historical context such as things people do and which influence them, for instance, knowledge, behaviour, suffering, nature etc. Furthermore, statements of a subject or topic fall into this category. Examples are "philately", "the Nazi past", "the Stettin Question", "the establishment and history of the 'Day of the Stamp'", or "the biography [of a person]".

Figure 12 shows the total amount of occurrences for each primary category of given entities. The importance of actors is evident with 609 distinct occurrences; that is, the sum of particular persons and groups and type of persons and groups, which amount to 37% of all given entities. Times (specific time-spans and historical periods) and events (particular events and activities as well as types of these) are the second most important given entities, with 284 (17%) and 281 (17%) combined occurrences respectively.

Equally relevant are particular documents and types of these with 207 (13%) and particular places and types of such with 168 (10%) occurrences. Other entities (general contexts, identifiers and things) are provided as given entities only in few cases.

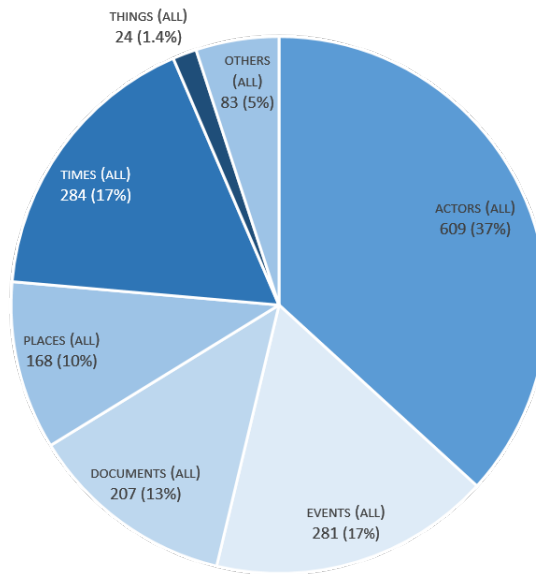


Figure 12 – Groups of given entities per total sample (n=1656).

Figure 13 further details the sub-categories of the three primary ones, *Actors*, *Events* and *Times* for the whole sample.<sup>60</sup> Within the Actors category, particular groups with 233 (38%) and particular persons 228 (37%) are the most relevant occurrences. In the case of Events, particular events with only 2 cases are close to irrelevant (0.7%) while types of events with 63 cases are more important (22%). The remainder of occurrences in this category are distributed among particular activities with 114 (41%) and types of activities with 102 (36%) counts. Finally, 261 (92%) time-spans represent the largest group in the Times category with only 23 particular periods (8.1%) and “no type” of period.

<sup>60</sup> Note that all percentage values pertain to the respective primary category; that is, to Actor (blue), Event (green), and Time (yellow).

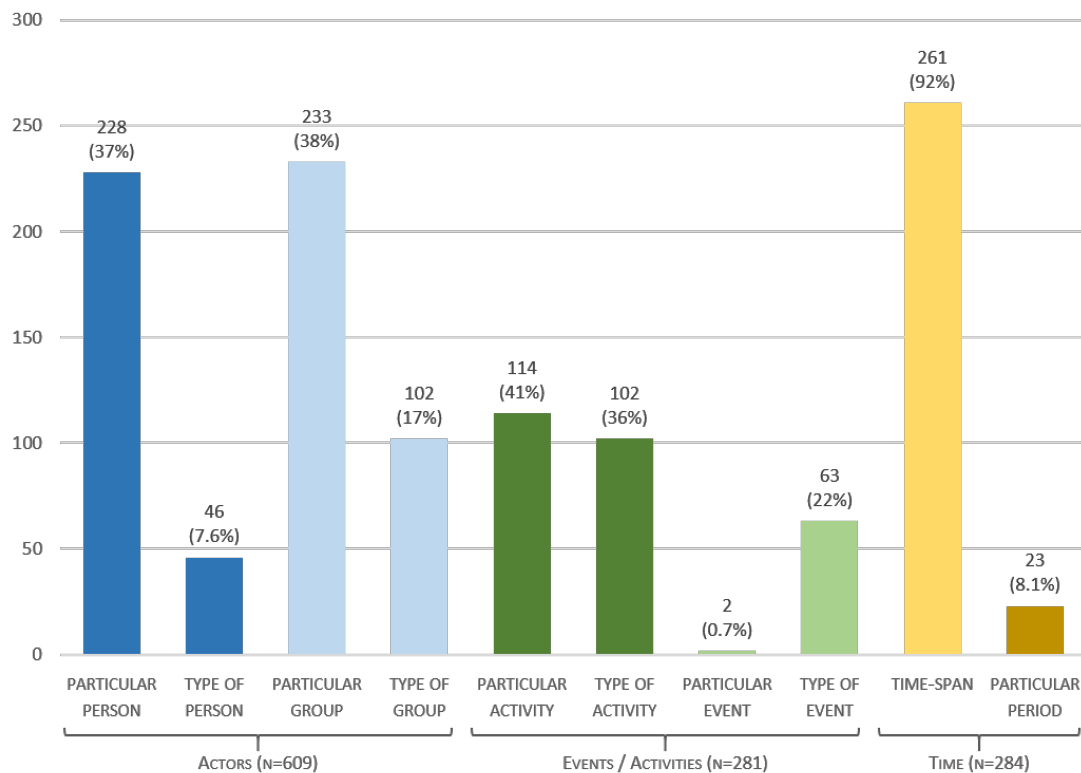


Figure 13 – Detailed listing of the three largest groups of given entities per total sample (percentages given per primary category).

Figure 14 shows the primary categories of given entities per sample. The relative distribution per sample diverges significantly only in the case of actors where the sample from the German Federal Archives exhibits close to twice as many occurrences than the sample from the National Archives of Norway. This difference, which is not shown in Figure 14, results primarily from a higher number of given particular groups in the case of the German Federal Archives sample with 204 occurrences equaling 41% of actors, versus only 29 occurrences in the case of the National Archives of Norway which represent only 25% of all actors.

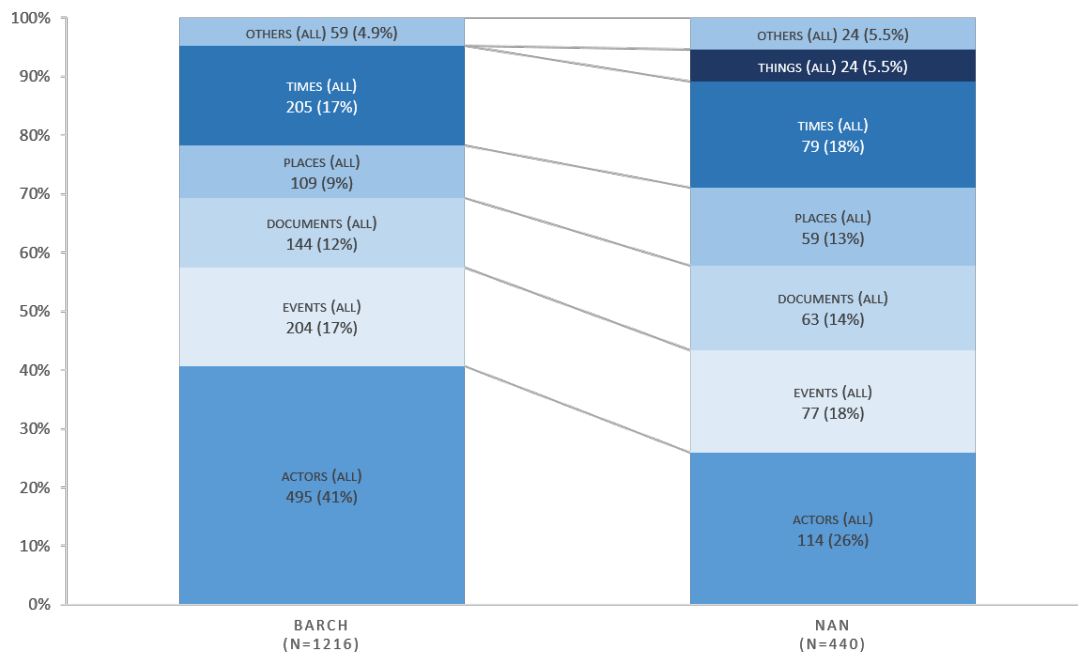


Figure 14 – Groups of given entities per individual sample.

The reason for this might be that the collected inquiries from the German Federal Archives sample appear to be more substantial in terms of the descriptive parts contextualizing their interest. Further, the users could be more proficient regarding which information is potentially relevant in the context of archival inquiries. This hypothesis might be further supported by the fact that, within the category of documents, the relative number of particular documents is significantly higher in the German Federal Archives sample – 84 occurrences which represent 58% of all documents versus 16 occurrences representing 25% of all documents in the National Archives of Norway sample – while, at the same time, the relative number of given document types is higher in the National Archives of Norway sample: 60 (42%) counts versus 47 (75%).

Furthermore, overall in the German Federal Archives sample, 71% of all given entities are particulars while in the National Archives of Norway sample this is the case for only 58%. The users in the German Federal Archives sample might happen to be more knowledgeable or experienced and therefore able to provide more specific information in their inquiries. The second apparent difference between the given entities in the two samples is the absence of *Things* from the German Federal Archives sample.

### 1.2.2 Wanted Entities

The types of questions already indicate whether the user primarily seeks archival or non-archival materials or factual information. The wanted entity therefore represents the entity which is primarily referenced by either the wanted archival or non-archival materials or by the wanted facts. The wanted entity indicates the perceived immediate ontological interest of the inquiry but not necessarily its general research interest. In this respect, the wanted entity is a compound statement derived from the type of question stating whether resources of any kind, resources of a specific type, a specific item, or factual information is wanted, as well as the entity *primarily* referenced by the wanted resource or fact.

For example, if the type of question is “specific item” and the wanted entity is “person relation” then the user is seeking one or more specific archival or non-archival documents that refer to a person and their direct or indirect relationship(s) with other actors; for example, the membership of the person in a political party. Note that these wanted entities are of course aggregations of more specific wanted entities that were subsequently subsumed into a shared and more abstract yet also more inclusive category during the analysis.

The referenced wanted entities are grouped into seven principal categories, which emerged iteratively from the analysis. They comprise *Actors* (including persons and groups) *Activities / Events* (including unintentional events and activities), *Documents, Things, Places, and General Topics*. Each category contains one or more sub-categories which specifically denote the nature or context of the wanted entity.

Table 2 – Wanted Entities

<b>Actor</b>	<b>Description</b>
<i>person or group</i>	A broad or unspecified amount of any kind of (biographical) factual information or documents either <i>about, by, or generally related</i> to a person or group. A typical principal example is a request for all information about the life and fate of a particular person. If the interest of the inquiry can be determined more specifically then one of the following sub-types is applied.
<i>- by person or by group</i>	In few cases the question specifically and exclusively asks for documents <i>by</i> an actor as the creator or keeper of those documents.



<p>- <i>person identity</i> or <i>group identity</i></p>	<p>Facts or documents related to an actor focusing on <i>who</i> performed or participated in an activity, the <i>existence</i> of an actor, the <i>name</i> or <i>identification</i> of an actor, including, as a special case, the responsibility or <i>mandate</i> of an actor. For example, the designation of a German police battalion in Norway during the Second World War, the proper name of a particular person, the name of a particular surveillance agency in the Weimar Republic, who negotiated, which battalion was in Norway, which organization had been responsible in the GDR for organizing journeys abroad.</p>
<p>- <i>person activity</i> or <i>group activity</i></p>	<p>Facts or documents related to an actor in relation to an activity focusing on <i>what</i> an actor did, or <i>if, how, and why</i> an actor did something; the behaviour and role of an actor, or the motivation and aims of an actor in relation to an activity, or the personal contribution of an actor to an activity with multiple participants. For example, whether an actor did or did not conduct or participate in an activity; whether a person has lived in a country or has been at a particular place, applying for citizenship, having participated in a demobilization or negotiations, having performed guard duty, whether it is true that Adenauer sent his son to Bordeaux to act as a defender, whether the church intervened, about the actions of a police battalion; How did they get to the UK? Why did so many desert? How did the state act regarding criminal prosecution? In what way did the FRG stir the investigations? Why did no chancellor visit Oradour? What was the position of the SED towards the ideological changes? How did a specific organization make decisions?</p>
<p>- <i>person perception</i> or <i>group perception</i></p>	<p>Facts and documents related to an actor and his or her perception, reception, assessment, attitude, or judgement. For example, how did authorities in the GDR perceive UN politics?</p>
<p>- <i>person relation</i> or <i>group relation</i></p>	<p>Facts and documents related to various kinds of direct or indirect relationships between actors such as family and membership relationships, including meetings. Examples are questions regarding whether two persons knew each other or whether a person was the member of a political party, held a specific citizenship, a particular nationality, or belonged to a particular district command.</p>
<p>- <i>person place</i> or <i>group place</i></p>	<p>Facts and documents related to an actor and a place. For example, the place where a particular police garrison was stationed in Norway during the Second World War, or the places where a particular police battalion was in Norway during the Second World War.</p>

<b>Event</b>	<b>Description</b>
<i>event</i>	Facts or documents generally related to an event such as an aircraft accident.
<i>activity</i>	Facts or documents generally related to an activity such as the minutes of the parliamentary sessions of the Reichstag.
- <i>activity role</i>	Facts or documents related to an activity and whether and how that activity occurred; that is, its principle existence or role, ramifications, or general history. For example, whether a criminal case initiated interventions, the role of the UNO policy in the state apparatus of the GDR, whether a prosecution was a topic in political relations between the GDR and FRG, the role of social policy in the propaganda of the SED, the history and establishment of the 'Day of the Stamp', whether there have been talks about organizing a meeting, or whether a journey has been discussed.
- <i>activity date</i>	<i>When</i> an activity occurred: Facts or documents related to an activity and a date such as when a medal was awarded, or a date of marriage.
- <i>activity place</i>	<i>Where</i> an activity occurred: Facts or documents related to an activity and a place, such as where an old Bible was printed.
<b>Document</b>	<b>Description</b>
<i>document</i>	Facts or documents related to various bibliographical information such as the number of existing copies, pointers to secondary literature or finding aids describing archival holdings. The Wanted document serves as a general and inclusive category in cases where the entity referred to or particular interest cannot be reasonably determined or does not fit any of the other categories.
- <i>document identity</i>	The Wanted comprises information on whether a particular call number is correct, or two holdings are identical, whether the type of document is a biography or report; it also includes questions related to the form of a document such as whether the document has been digitized.
- <i>document extent</i>	The Wanted comprises information on the extent of a file, series, holding, collection etc. such as how many files the holding contains, or how many pages are in the file.
- <i>document provenance</i>	The Wanted comprises information on the past and current custody or holding of documents such as whether the files of the district committees of the Kulturbund have been coherently delivered to the responsible town archives, whether the archive assumed custody of films, or where a particular archival collection is being administered.

- <i>document content</i>	The Wanted comprises information related to the contents of an archive, collection, aggregation of items or single items such as a general content description, whether a specific person is mentioned in a text, whether specific features such as personal marks appear, or what specific markings in a text mean.
<b>Thing</b>	<b>Description</b>
<i>thing</i>	Facts or documents related to a thing such as a building or features of a building, ships, submarines, or aircraft.
<b>Place</b>	<b>Description</b>
<i>place</i>	Facts or documents related to a place such as a map of a particular city, a map of a drainage system and plant of a particular place.
<b>General Topic</b>	<b>Description</b>
<i>general topic</i>	Facts or documents related to a general topic such as "correspondence for armament and war economy", "limits of growth", "zero growth", "ecology", or "milk analysis project".

Figure 15 shows the distribution of wanted entities among the primary categories for the whole sample. The primary category *Actors* leads clearly with a count of 292 which equals 61% of all wanted entities in the whole sample. Also significant are *Activities / Events* with 122 occurrences (26%), of which only two were unintentional events, and further *Documents* with 43 occurrences (9%).

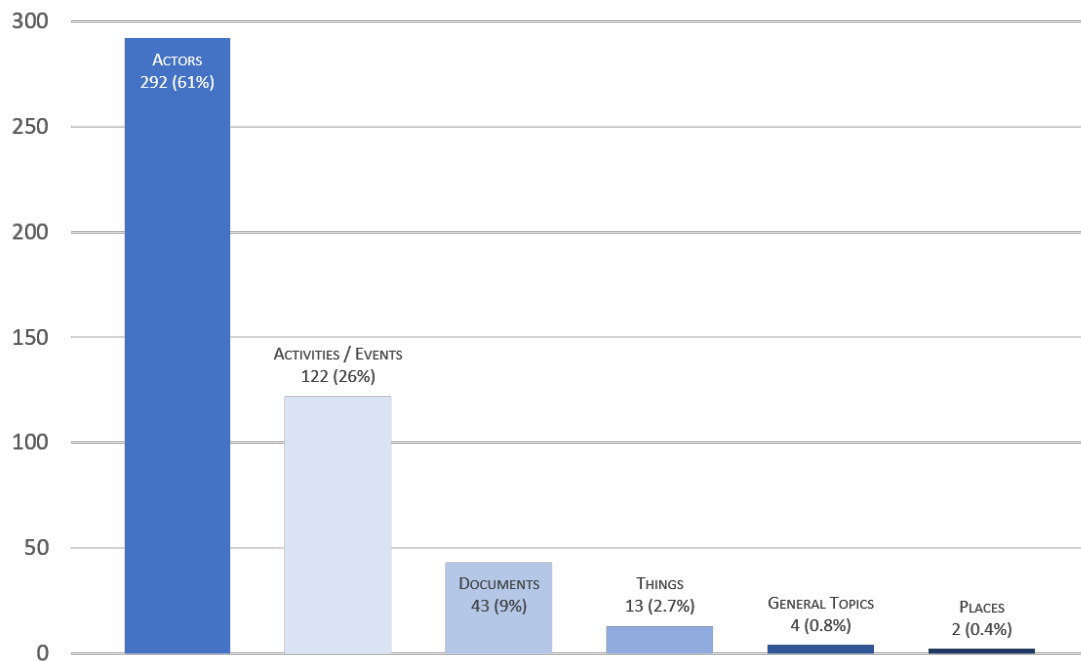


Figure 15 – Primary categories of wanted entities.

Figure 16 shows the distribution of the wanted entities per sample. The relative distribution exhibits no significant deviations, with the exception of the category *Things*, which only appears in the National Archives of Norway sample, and a slightly higher relative number of *Activities* in the German Federal Archives sample. Further, *Events* and *Activities* are displayed distinctly in this figure in order to show that only two unintentional events occur as wanted entities in the National Archives of Norway sample.

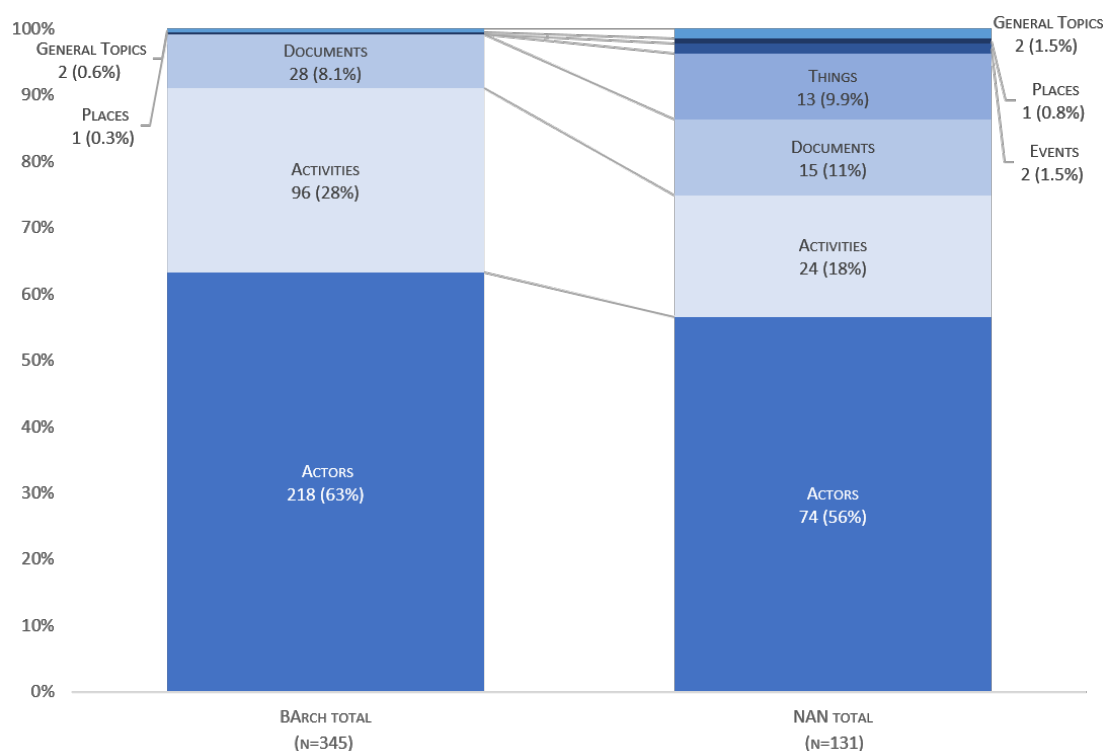


Figure 16 – Primary categories of wanted entities per sample.

The primary category *Actors* can be further divided into groups and persons. The numbers for both entities correspond in both samples: persons occur 138 times in the German Federal Archives sample, which amounts to 63% within the *Actors* category, while in the National Archives of Norway sample only 47 persons occur as wanted entities (64% within the same category). Groups appear 80 times (37%) in the German Federal Archives sample and 27 times in the National Archives of Norway sample.

Figure 17 further differentiates between the various sub-groups of the primary category *Actors*.<sup>61</sup> Inquiries generally regarding actors constitute by far the largest group in both samples followed by actors in the context of activities.

<sup>61</sup> Note that *persons* and *groups* have also been differentiated during the analysis but are not displayed in the figure.

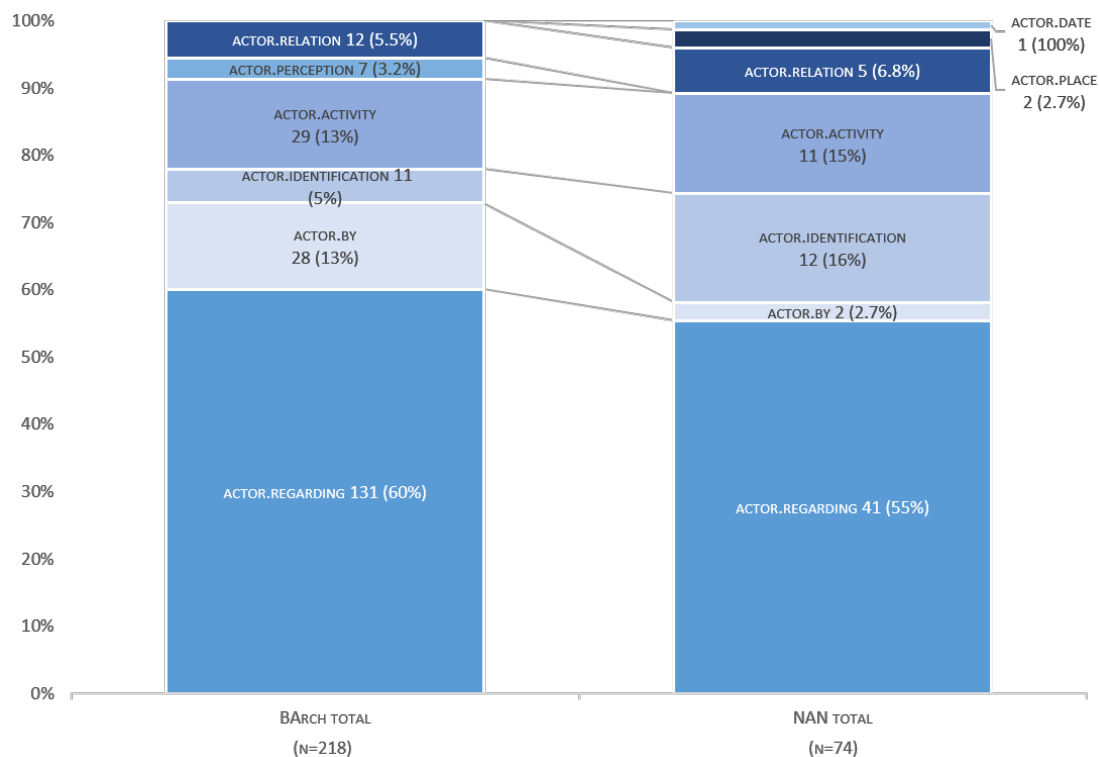


Figure 17 – Details of the wanted entities group Actors.

Archival materials by an actor are more important in the German Federal Archives sample while inquiries regarding aspects of identification are more important in the National Archives of Norway sample. Questions about the perceptions of actors only occur in the German Federal Archives sample while relationships between actors are present in both. Actors, dates and places are insignificant in both samples.

The second largest primary category of wanted entities, *Activities*, is shown in Figure 18 together with its sub-groups. Inquiries generally interested in activities constitute the largest group in both samples with 73 occurrences (76%) in the BArch sample and 17 occurrences in the National Archives of Norway sample (79%).

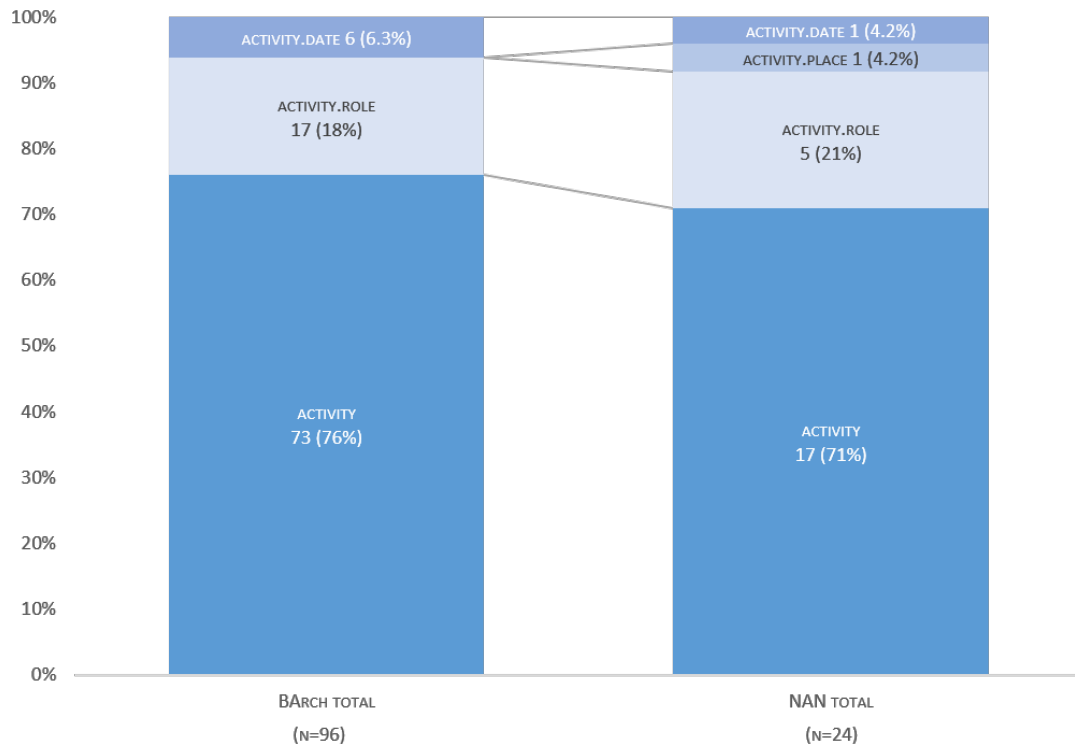


Figure 18 – Details of the wanted entities group Activities.

The role of activities (activity.role) and, to a greater extent, of activities and dates (activity.date) are much less important. Only one inquiry in the National Archives of Norway sample concerned an activity in the context of a place (activity.date).

Finally, the distribution of the primary categories of wanted entities among the two main categories of the question types resource discovery and fact-finding is shown in Figure 19. The wanted entities are slightly more diverse in the case of resource discovery questions, while actors and documents appear to be more important in fact-finding questions.

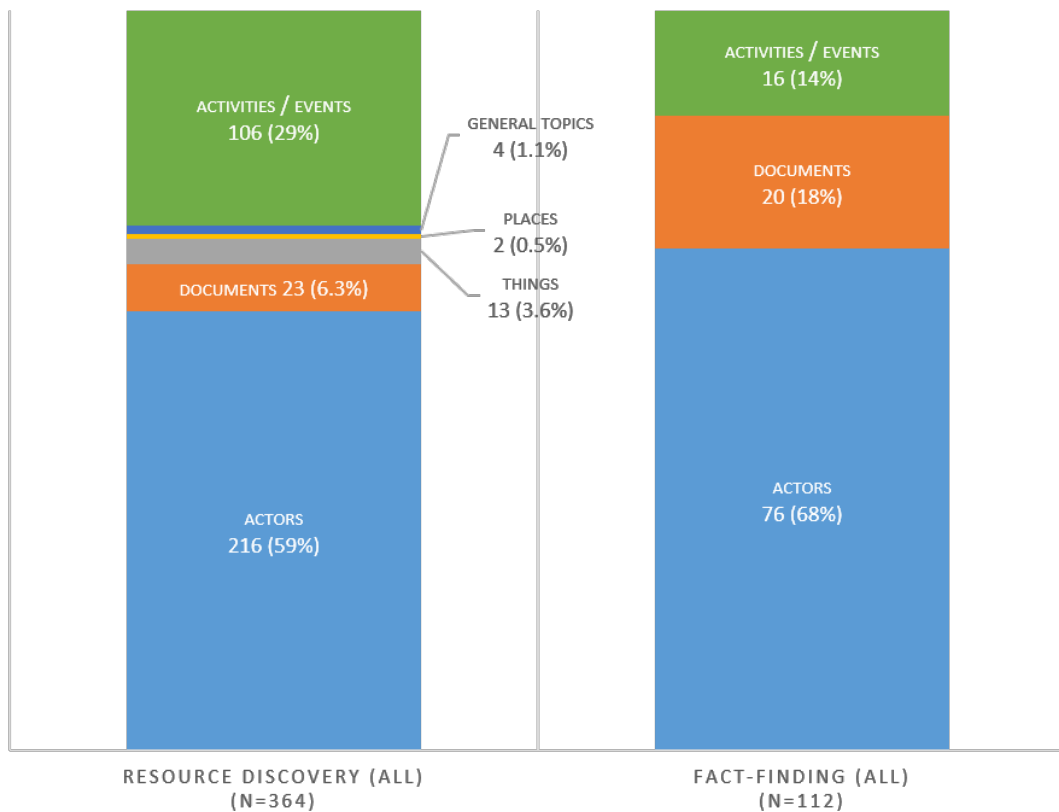


Figure 19 – Primary categories of wanted entities per primary categories of type of questions.

As discussed in the previous chapter, the wanted entities cannot be easily compared to the Wanted in Duff and Johnson (2001) since both are based on different conceptualizations. The interest of users in actors is apparent from the figures discussed. Interestingly, not only individual persons but also groups are relevant. In this context, the discovery that activities are also significant interests of inquiries, which contradicts the findings by Duff and Johnson (2001), is important since activities are acts carried out by persons and groups. This discovery thus further underlines the relevance of actors. Unintentional events, on the other hand, are insignificant.

The various identified sub-types of wanted entities indicate that the interests are more diverse than the primary categories may suggest and can provide initial indications for the further interpretation and formalization of the interest of the inquiries.

Even though the wanted entities are more meaningful and expressive than the Wanted described by Duff and Johnson (2001), they constitute only a preliminary understanding of the users' interests and the relevant subject matter to which these interests pertain. The wanted



entities are only the first step towards a more thorough and explicit representation of these needs. As the initial building blocks of the ontological model, the given entities and especially the wanted entities need to be further analyzed and investigated with respect to which relevant relationships exist between them.

One important intermediate step during this continuing interpretative process is to determine how directly the further interest of an inquiry can be translated into an ontological representation; that is, material facts (IV:3.1) adequate to the archival and historical epistemological framework (IV:2).

### 1.3 Interest of Inquiries

The CRM represents a world of *material facts* (IV:3.2), since it is empirically based on the analysis of metadata standards, which in turn primarily represent such material facts. The contents of archival descriptions also mostly consist of material facts; that is, typically do not explicitly describe psychological, collective or statistical facts. The archival and historical inquiries posed to an archive may, on the other hand, very well be interested in such kinds of statements and facts, for example the motivations or intentions of an actor, as discussed in the previous chapter (IV:2.1).

Such inner states of an actor, however, cannot be reasonably expected to be explicitly documented in an archival description. In other words: There are observable facts; that is, material facts, and facts that are not directly observable, of a psychological or social nature. Only the former can be reasonably expected to feature in archival descriptions. In order to find documents in the archive which may provide insights on such inner states or other states that cannot be directly observed, a question would need to be broken down into secondary or auxiliary questions that pertain to a framework of material facts describing and pointing to appropriate documents.

An important step during the interpretative analysis is therefore the categorization of questions according to whether the initial question can be directly answered by a potential query pattern. In other words, whether or not potential adequate documents or facts for answering the questions can be immediately described by material facts.

At the same time, the categorization represents the specificity of the interest of the inquiries posed to an archive and has evolved from the iterative interpretative analysis of both these inquiries and the ontological modelling of their interest. Three categories have been identified: *material fact* questions, *psychological* questions, and *collective statistical* questions.

#### Material Facts

*Material fact* questions seek a directly observable *relationship* between two entities. A question such as: “Is it true that Adenauer sent his son to Bordeaux?” or, in more general terms, “Has a specific actor been at a certain place?”, would be a *material fact* question because the adequate

answer to this question can be found in a statement such as: “Person – has been at – Place”.

Questions such as “Did Fritz work for the Stasi?”<sup>62</sup> or, more generally, “Did a specific actor work for a specific institution?”, would also be a *material fact* question even though this historical relationship would not have been “observable” by arbitrary actors. However, the relationship is in principle observable and potentially documented in archival records.

All inquiries need to be represented as material facts in order to enable an archival information system to serve the interest of an inquiry. Most questions, as will be shown, fall into this category, and their translation into probable and adequate answers proved to be relatively straightforward.

Questions which cannot be answered directly with material facts need to be interpreted further in order to break them down into secondary questions that can be answered indirectly based on material facts leading to appropriate documents or facts. Such questions can be differentiated into *psychological* inquiries, which primarily consider the inner states of actors, and *collective statistical* inquiries, which primarily focus on the actions of groups and particular regularities of behaviour.

The process of deducing relevant secondary questions and breaking down original or initial questions into more specific sub-questions is pivotal to the methodology of historical inquiry. *Psychological* and *collective statistical* questions need to be investigated and conclusions substantiated by evidence based on primary facts, which again pertain to material-fact statements. In the archive in particular, archive users as well as archivists explicitly or implicitly apply such techniques when conducting research for documents or factual information.

### Psychological Questions

*Psychological* questions are about *inner or mental states* of an actor, a person or group, and relationships which are *not immediately observable*. The comparison of a statement or action of an agent with the inner state of mind or intention of said agent, for instance, are of a psychological nature. Further, such questions typically inquire as to reasons, intentions and motivation of actors with regard to activities in which they have participated.

The question, “Why has no German chancellor or president visited Oradour?”, for example, does not ask whether a person has or has not been at a certain place, but inquires as to the reasons and motivation for the decision to visit or – as in this case – to refrain from doing so. These kinds of questions typically demand a higher level of interpretation in order to find material facts providing adequate and probable answers to the question.

Another example is the perception or appraisal of activities by an actor. For example, the question, “How did the responsible authorities in the GDR see the policy of the GDR towards the UN?” refers to the psychological and inner state of an abstract governmental body or, in other words, how agents may have perceived a specific policy.

---

<sup>62</sup> “Stasi” is the colloquial abbreviation for *Ministerium für Staatssicherheit*, the Ministry for State Security of the German Democratic Republic, acting as an intelligence and secret police agency.

## Collective Statistical Questions

*Collective statistical* questions focus on how groups have acted, which may entail questions about their reasons for doing so, regularities and frequencies of collective or individual behaviour, of certain types of events, or other phenomena. These phenomena take on a relevant statistical form that cannot be considered random or coincidental and that thus has consequences for society or other actors. Questions of a *collective statistical* nature are not always clearly distinguishable from questions of a psychological nature. Some questions may fall into both categories or constitute the basis for answering *psychological* questions.

The question “Was Fritz an avid informer of the Stasi?” is an example of a *collective statistical* question. The adjective “avid” indicates that the interest is not in the fact as to whether Fritz worked for the Stasi or not but rather in the intensity with which this person made a contribution or in the quality of the delivered contents.

The intensity of an activity could be assessed based on the frequency with which specific types of activity occurred or were carried out by the person in question over a certain time-span. For example, the inquirer could investigate how many reports were delivered by Fritz to the Stasi within a certain time-frame and which can be found in an archive. Finding documents created by a particular person and delivered to a particular group then is a *material fact* question and therefore suitable for an archival information system.<sup>63</sup>

Based on such documents, the inquirer could then begin to assess whether or not the person in question could indeed be characterized as an *avid* informer, based, for example, on the dates of the documents. Whether the contents exhibit high quality in terms of, for example, comprehensiveness is a matter which can only be judged by the researcher. The archive only provides the empirical source material in order to investigate such questions.

Furthermore, the type of initial question as to whether or not Fritz was an avid informer of the Stasi could be categorized either as a *fact-finding* question or even a *research* question during the linguistic analysis. As a *fact-finding* question, the user would expect the archivist to evaluate the eagerness of Fritz’s contribution and possibly to provide some examples or evidence for the verdict. For this reason, fact-finding questions are also assumed to require documents as evidence for a fact or as the source for finding the fact.

The question as to whether “the wives’ joining of a political party coincided with the political influence of their husbands or if the women had been politically engaged before they met their future spouses” also belongs to the type *collective statistical* because in order to provide an answer two dates need to be compared: the date of joining a political party, if any, and the date of marriage.

Of course, *collective statistical* and *psychological* questions cannot always be precisely assigned or strictly delimited. For example, a question such as “What was the course of action of the

---

<sup>63</sup> Of course, there would be various practical and legal obstacles to identifying a particular person as an informant of the Stasi and to finding documents which have indeed been created by that person, such as being listed only under an operative name. Such legal issues, however, lie beyond the scope of this study and the ontological model.

Greens while trying to achieve their goals in their *Deutschlandpolitik* regarding the GDR?" is primarily a collective statistical question, yet reasonable secondary questions may very well ask for the personal motivation of leading figures among the Greens, thus rather focusing on psychological aspects.<sup>64</sup> The categorization therefore should not be taken as absolute but rather as an approximation of the general level of interest expressed in the single inquiries.

### Indirections

Inquiries may not only suggest reasonable secondary questions but also exhibit more than one recognizable interest. Additional possible interpretations which may serve the interest of an inquiry are called *indirections*. This is especially, but not exclusively, the case with psychological and collective statistical questions which often also allow for various degrees of adequate inclusiveness or exclusiveness in terms of potential query patterns. Inquiries may therefore be served by more than one query pattern.

Questions may be answered more exclusively with a more specific query pattern, or more inclusively with a broader query pattern. Queries may "move up or down the ladder" in order to increase or decrease the result set. The query patterns represent only selected and the most relevant and instructive aspects of inquiries. Especially in cases where inquiries are very general or unspecific, for example research inquiries, or are of the type *psychological* or *collective statistical*, they entail several indirections.

These emerging different layers of interest; that is, of the epistemological interest of the questions, are reflected in the general patterns constituting the AKM which allow for movement from more specific and exclusive levels to more general and inclusive levels. This hierarchical organization of the general patterns, which will be discussed in greater detail in the following section, facilitates the representation of nearly all questions with respect to their interests.

Selected indirections will be discussed in the context of the examples provided with each general pattern and some will also be displayed in the diagrams accompanying the exemplary query patterns. However, it is neither the objective nor within the scope of this study to extrapolate all possible indirections resulting from the interpretation of the inquiries.

### Summary

Figure 20 shows the relative distribution of inquiries among the three levels of interest. In the whole sample, consisting of 476 substantial questions in total, 399 (84%) are material fact questions. The remaining 77 questions (16%) consist of 24 (5%) psychological questions and 53 (11%) collective statistical questions. A comparison of the relative distribution of the three levels of interest in the two samples of the German Federal Archives and National Archive of Norway reveals that only psychological questions differ significantly. The high number of fact-finding questions in the NAN sample may explain this, since these tend to be material fact questions.

---

<sup>64</sup> Such *indirections* during the interpretation process are discussed below.

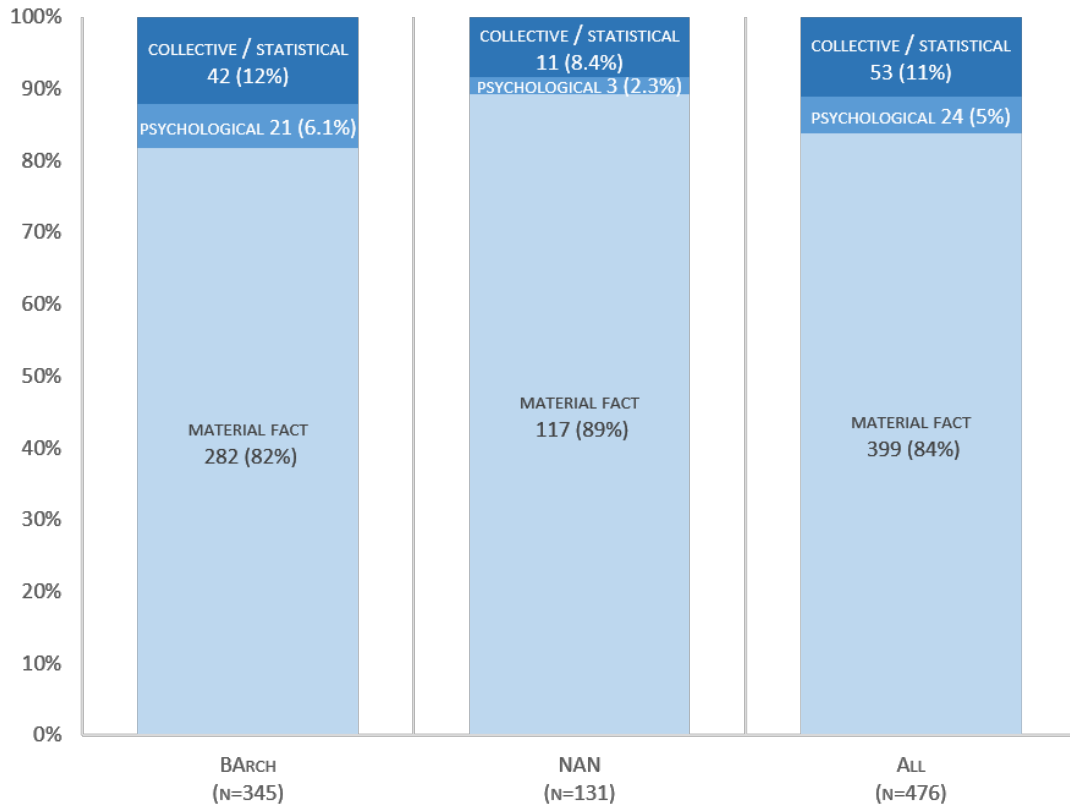


Figure 20 – The level of interest of inquiries.

One conclusion which can be drawn from these figures is that most users consciously or subconsciously produce many of their inquiries as fitting to an archival context. In other words, the user already conceptualizes and formulates questions as more-or-less appropriate potential queries to an archival information system. Many *material fact* questions demand less interpretation in terms of what the user's interest is and how this interest can be adequately represented in the form of general patterns. The challenge then is to interpretatively break down psychological and collective statistical questions into appropriate material facts which adequately represent the perceived interest of the questions. As previously discussed, the material facts constitute the general patterns and are essentially the result of reducing complexity to an adequate ontological level of abstraction and simplification.

The interpretative analysis further produced and, at the same time, evolved around an *ontological core framework* which formalizes and describes pivotal entity types and relationships pertaining to the essential interests of the inquiries. This framework will be discussed in the following.

## 1.4 Ontological Core Framework

During the interpretative analysis of the single questions and their associated contextual information an ontological core framework emerged which represents the most dominant entity types and relationships as identified in the analyzed inquiries. Together these entities form an ontological representation of the historical reality which constitutes the domain of discourse expressed in the inquiries and which is pertinent to their interests.

Figure 21 shows this ontological core framework that has guided the ontological analysis of the user inquiries. The diagram formalizes the most dominant contexts and core entities of the shared domain of discourse of the inquiries; in other words, the most relevant, top-level ontological entities constituting the subject matter to which the interest of the inquiries pertains.

To provide a short overview of the figure, which will be discussed in more detail further below, we can begin by stating that activities are carried out by actors; that is, by individuals and groups. These activities may be the subject of information objects (written or oral accounts), which in turn have been created during another activity, often a designated act of documentation during which an activity has been observed. In this case, the information object can be understood as an actual document of the observed activity.

Activities are further characterized and realized by the fact of their execution. The execution exhibits a particular structure revealing how the activity has transpired. The document about the activity then is evidence of its execution, although not proof, and may describe the structure of the act; in other words, recount what happened or is said to have happened.

The actor carrying out an activity does so based on a will; that is, the actor has the will to perform the activity. The will is a temporal entity and may be spontaneous, reducing the likelihood of evidence of this will existing. However, expressions of will may be recorded in symbolic form such as in a recorded statement of intent or an announcement that an activity will be undertaken. The “unwilling” participation of an actor in an event or the mere presence of an actor at an event or activity is not covered at this point of the analysis. If an activity is known to have been carried out, then there must have been the will for its execution.

The willful undertaking of an activity often pertains to a plan or, more precisely, to a prospective plan structure: how things should be done or what should be achieved and how. Similar to act structures, which may be described in information objects, plan structures too may find symbolic expression as an information object or as part of such. By comparing records of plan structures with records of act structures, historians are able to obtain valuable evidence and factual information for establishing historical facts.

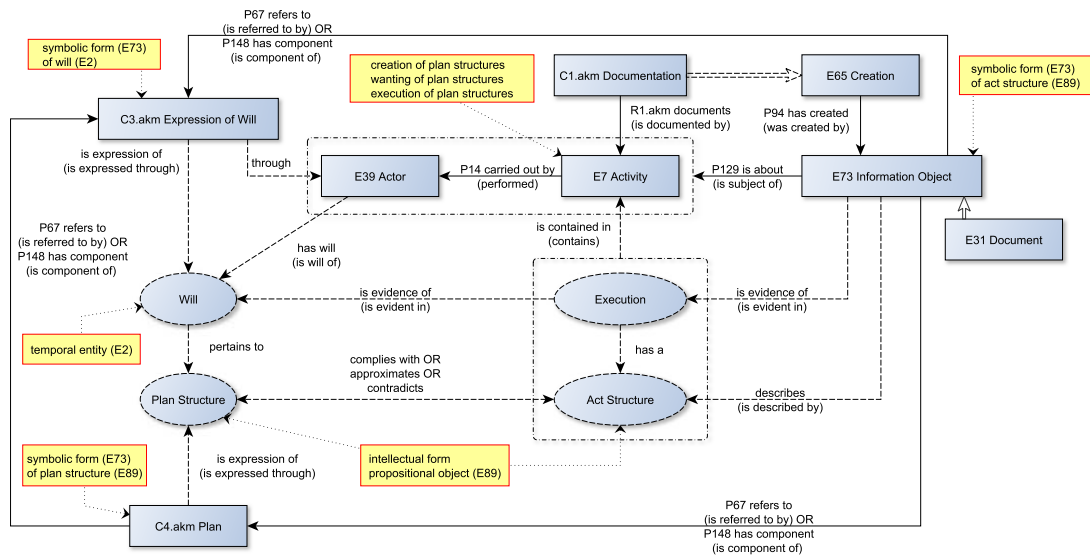


Figure 21 – Ontological core framework of the Archival Knowledge Model.

The notation used in the diagram has been introduced in the previous chapter (IV:3.3). The current diagram utilizes two additional notational elements: *dotted single ellipses* and *dotted single arrows*. Both of these represent ontological entities; that is, entity types and relationships respectively, which are not explicitly formalized in the AKM but constitute pivotal entities of the domain of discourse as described by the inquiries and the epistemological framework delineated in the previous chapter.

These ontological entities are not immediately observable, such as the will of an actor or the structure of an act, and are not likely to be directly or explicitly documented in an archival information system. They are, however, important ontological entities that constitute crucial background knowledge with which to interpret the interest of the inquiries, and which find their expression in classes and relationships of the AKM. This will be discussed in greater detail in the context of the general patterns (V:2).

The other ontological entities represented in Figure 21 are already explicitly formalized in the AKM as classes and relationships. The most dominant wanted entities – actor, activity and information object (documents) – are easily recognizable. Other entities, and especially the relationships between these entities, are the result of the extensive and iterative interpretation of the inquiries.

By representing the dominant entity types and relationships in the interests of inquiries, the framework also represents the historical point of reference for the interpretation of the inquiries. The framework guided the successive integration of the various phenomena subsequently identified in the inquiries into a coherent model of interrelated ontological entities. In this

regard, the framework constitutes a kind of intersection between the user inquiries and the AKM.

In other words: The diagram can be understood as a stable snapshot of the iterative and recursive interpretation and modelling process, describing a state between the already aggregated and abstracted interpretative results and their ontological formalization.

The identified core entities formed a consistent, yet intermediate core framework and served as the foundation for the more detailed reinterpretation of the observations until they appeared as a consistent and stable arrangement of generalization and specialization as will be described in the section “General Patterns” (V:2). As such, the ontological framework is clearly the result of the interpretative analysis and, due to its efficacy, only appears as its logical prerequisite at the end of the analysis.

The AKM and its general patterns derive from this ontological core framework. The general patterns introduced in the following section essentially supplement and extend this ontological core framework with additional, specializing or generalizing classes and relationships. Each general pattern specifies one particular segment of this framework as will be discussed in the following section: “General Patterns” (V:2).

The remainder of this sub-section will describe the essential notions of the entities represented in the diagram as well as provide a selective high-level characterization of the represented domain of discourse. All entities will be discussed in greater detail in the context of the various general patterns.

**Information Objects** An *information object* carries information in a symbolic form, such as letters and words, but also graphics such as drawings or carvings. The information objects represented here, the plan, the expression of will, and the document, may reside on the same physical carrier or be contained within one another.

They also have different notions; for example, the expression of a plan structure projects into the future, such as an agenda or draft, while an expression of act structure looks back into the past; for example a report or dossier, and an expression of will may also either project into the future, be an intention, for example resolutions or statements, or may be an opinion, thus focusing on the moment.

The will and plan belong to an *intentional level (Absichtsebene)*. Both find their symbolic expression in information objects, typically documents. The act or activity and the actor are part of the *historical reality*. Both realms are documented and thus accessible via information objects. The archive is the institution that provides us with these documents. These documents are the point where this ontological representation of the historical reality directly intersects with that of the *archival domain* of record keeping.

These information objects are the only trace and evidence we have of *plan, will* and *act* in the archive. The will becomes only evident in information objects. These are the preferred evidence by historians of the potential existence of an actor or actors and what they did, wanted and



planned; all these past occurrences are only indirectly accessible through the traces, most of which are in written form, stored in the archive.<sup>65</sup>

**Plan and Will** The word *plan* denotes two meanings: (1) to have a plan for something, a schema or description of a procedure or to arrive at a particular state or form of something, a plan structure, (2) to plan to do something: the will to realize or act upon a plan structure or to perform an action. These two notions of plans are distinctly represented in the model.

The *plan structure* may have three types: (1) the plan as the description of *appearance*, a plan of a building, for example; (2) the plan as the description of a *procedure*, such as a *sequence* of prescribed steps, a legal regulation, a law or mandate; and (3) a plan for *achieving* some condition or state of affairs, such as a political agenda. Specific examples include resolutions, preparatory material, death sentences, city plans, etc.

The intellectual plan structure finds its symbolic expression in an information object which is an entity explicitly represented in the AKM, labelled as *C4<sub>akm</sub> Plan*. Any such plan or plan structure necessarily projects into the future with regard to activities based on a plan structure.

If a plan or plan structure is executed then there must have been a *will* to do so. The *will* is a temporal entity and refers to psychological facts as mental or inner states of a person or group, their intentions and opinions. They can have a will to do something, to enact a plan for example. This psychological world is not formally represented in the AKM. The will becomes evident only indirectly in the archive through a documented expression of it.

Separating the will from the plan structure allows for a distinct representation of how drafts of plans are created, and enacting, in other words '*wanting*', these plans, such as by parliamentary vote, until the plan is no longer desirable, such as with the abolition or annulment of a law or order.

Similar to the plan structure, the will finds its symbolic expression in information objects and is an entity explicitly represented in the AKM, labelled *C3<sub>akm</sub> Expression of Will*. An expression of will is typically part of another information object such as a plan or document. For example, a Declaration of Independence is a plan and contains the will to be independent.

The will as a temporal entity forms and evolves over time as an inner state and then finds singular expressions of particular states of itself in different forms, such as in speech or writing or in characteristic actions. Evidence of such acts of communication as driven by wills are found in the archive, typically indirectly as part of documents that refer to the will in some manner, by containing a signature for the enactment of a law, for example, or by recording the statement made during a speech proclaiming to do something.

The inner forming of a will is not observable and therefore can only be traced by finding and comparing records of such single acts of communication or statements referring to or reporting on particular states of the will. Comparing such statements allows us to draw conclusions on

---

<sup>65</sup> Contemporary witnesses are excluded from the discussion. They are living witnesses of the past and are as such also potential sources of evidence for historians.

how the will has evolved. By further comparing the stated will to records of activities and their plan structure, carried out as a consequence of the expression of will, the extent of its realization can be assessed by the historian.

To find out whether or not a specific will has actually existed is the task of the historian by comparing plan structures with how actors have conducted activities; that is, with records of these activities. The historian then may find that the two propositional entities either comply with, approximate or contradict each other, and then draw further conclusions based on these basic assessments.

**Act and Actor** The *act* is an activity consciously carried out by an actor. In contrast, unintentional events can be understood as “non-activities” occurring without the intentional participation of or immediate causation by humans. Every activity consists of the fact of its *execution* and of an *act structure* which refers to how the activity has occurred. The class *E7 Activity* explicitly represents the act in the AKM and implicitly contains the ontological entities’ execution and act structure. Furthermore, the act cannot be without an actor since it is dependent upon an actor in order to be carried out. The class *E39 Actor* explicitly represents the actor (persons and groups) in the AKM.

Both the actor and activity are historical entities that are no longer manifest.<sup>66</sup> Only information objects may provide direct or indirect evidence of their existence: a document may be evidence of the execution of an act structure by containing an account of said act structure or even only stating the execution or occurrence of an act. In this case, the information object (*E73*) is about an activity (*E7*) and thus indirectly also a testimony of the existence of a specific actor, who may be known or unknown, but who must have carried out the act. In few cases, an information object may be directly and exclusively about an actor, for example a description of the features of a person or group.

These information objects exist in the archive because the activities have been documented in a way that leads to the creation of said information objects. Documentation is itself an intentional activity only in that it necessarily creates an account of some other activity. Both activities are explicitly represented in the AKM as the class *E65 Creation* and its sub-class *C1<sub>akm</sub> Documentation*.

Inquiries request any or specific information objects or facts – indicated by the type of question – generally pertaining to the historical context represented by this ontological core framework. These information objects or facts may, for example, be about or refer to any entity in the framework such as activities, events, actors, plans and wills, or information objects themselves. The general patterns fully describe these relationships, which can be further distinguished into principal notions of “aboutness” of said information objects and facts, and the provenance of information objects.

---

<sup>66</sup> Contemporary witnesses that are still living, as previously mentioned, as well as such artefacts as tombs, bones, or mummies are not considered since the discussion focuses on the archival domain.

The ontological core framework described in the previous paragraphs not only introduces the most dominant classes and relationships forming the basic layer for the general patterns, but also a fundamental distinction between a context of provenance and a context of *aboutness*. This distinction forms the basis on which the interest of inquiries in terms of the wanted information objects can be represented, but also indirectly lays the groundwork for factual information, as will be discussed later on. The introduction and discussion of these two contexts in the following sub-section will lead on to the presentation of the general patterns.

## 1.5 Provenance and Aboutness

The previously introduced general ontological core framework formally describes a historical reality which constitutes a pivotal point of reference for the interpretation of the inquiries and the creation of the general patterns. On this basis, ten general patterns evolved during the ontological analysis and the subsequent ontological modelling process. These general patterns are shown in Table 3 and will be discussed in great detail in the following section (V:2). This subsection provides some background information on the general patterns.

When this core framework is related to the archival context, two principal types of general patterns become apparent: one type describes the provenance of things while the other describes the “aboutness” of these things.

During the ontological formalization of the results of the interpretative analysis, two principal kinds of historical context emerge by means of which the general patterns are characterized and intellectually grouped:

1. The general patterns belonging to the *provenance context* focus on the *historical context* from which the requested things have resulted, which includes acts of *creation* and *possession*.
2. The general patterns belonging to the *aboutness context* focus on the *historical context* to which the inquiry primarily refers, either directly or indirectly, via the requested documents.

In the case of *resource discovery* questions, two general patterns, one from each context, always represent the subject matter of the primary interest of the inquiry: one general pattern from the *provenance context* provides entities to describe creation, keeping, and taking notice of things, and one general pattern from the aboutness context specifies the historical reality to which the inquiry and the requested things refer. Selected elements from both patterns are instantiated with the given entities and in few cases with additional information derived from the interpretation of the inquiry. These then form a query pattern constituting a categorical and exemplary representation of potential queries adequate to retrieve relevant information objects or facts.

Factual questions may, in principle, inquire about any aspect of the provenance or aboutness context. Furthermore, since this study assumes a fundamental need for evidential materials in

order to provide and support the factual information requested by the user (IV:2.1), information objects potentially capable of providing such evidence are implicitly considered secondary objectives of the interest of the question. Factual questions are therefore also assigned two patterns, one from each context.

The general patterns grouped in both contexts differ in their level of specificity and are partly sub-patterns of each other. The patterns *Provenance* and *Aboutness* constitute the uppermost layer of the AKM in terms of its semantics. Both patterns will be discussed at greater length in the next section since these two patterns also pertain to all others on account of inheritance (IV:3.1).

<b>Provenance Context</b>	<b>Aboutness Context</b>
<i>Provenance</i>	<i>Aboutness</i>
<i>Documentation</i>	<i>Events</i>
<i>Self-Documentation</i>	<i>Plans</i>
<i>Correspondence</i>	<i>Actors</i>
	<i>Things</i>
	<i>Documents</i>

Table 3 – Provenance context and aboutness context general patterns.

The general patterns *Provenance*, *Documentation*, *Self-Documentation*, and *Correspondence* belong to the provenance context. The general patterns *Aboutness*, *Events*, *Plans*, *Actors*, *Things* and *Documents* belong to the aboutness context.

The general patterns belonging to the aboutness context indicate the general “aboutness” of the questions; for example, if a question is assigned to the *Actors* pattern it will usually be about one or more actors. In this regard, general patterns pertaining to the aboutness context indicate unspecified “aboutness”.

The general patterns belonging to the provenance context indicate the general context in which an information object has been created. The *Provenance* pattern indicates an unspecific context, either of creation or of the storing of information objects or physical things.

One could say that the general patterns place the given and wanted entities within relevant and adequate historical contexts representing the historical reality to which the primary interest of the inquiry refers.

The assignment of a general pattern from the provenance context conforms to questions regarding the most likely historical context of provenance from which the documents in question resulted. For example, a question about surveillance reports would be assigned to the *Documentation* pattern since such reports resulted from observing the activities of others. This sense is covered by the *Documentation* pattern.

The assignment of a general pattern from the aboutness context conforms to questions about the most recognizable historical context to which the question refers. For example, a question

inquiring into the membership of a person in a political party would be assigned to the *Actors* pattern since the question is generally about the actor. This sense is covered by the *Actors* pattern.

The particular combination of two general patterns is deemed best suited to represent a query pattern adequate to reasonably answer the most recognizable and primary interest of the question. In both cases, the most specific general pattern possible is chosen that would not, however, exclude potential queries of a more general nature.

While the general pattern itself already characterizes the interest of an inquiry, when instantiated in the context of query patterns a *primary entity of interest* can be determined that further indicates which particular entity is most relevant in each general pattern used. This goes for both the case of general patterns from the provenance context in terms of who created, produced, or kept an information object, and the case of general patterns from the aboutness context in terms of the specific aboutness of said information object. For each query pattern, therefore, a *primary entity of interest* is further provided in square brackets. This indicates one element within the respective general pattern which represents the most recognizable and relevant ontological entity indicating the aboutness and provenance of things or facts requested in an inquiry within the context of a particular general pattern.

In the case of general patterns pertaining to the aboutness context, the primary entity of interest denotes the type of entity the question in general or the requested information objects or facts are primarily about. In most cases, the primary entity of interest conforms to the wanted entity. For example, the assignment of the general pattern *Actors* describing the aboutness context with the primary entity of interest [*particular group*] would mean that the inquiry and request for information objects or facts are primarily about a particular group.

In the case of general patterns pertaining to the provenance context, the primary entity of interest is the given entity, which is the creator, producer or keeper of the requested information objects or physical objects. The rest of the given entities are put into selected “slots” of the general patterns, thus forming an exemplary query pattern. If an entity is unknown, then “any” is used as the primary entity of interest. For example, the assignment of the general pattern *Documentation* describing the provenance context with the primary entity of interest [*particular person*] would mean that the inquiry asks for information objects as the result of a documenting activity carried out by a particular person.

Assuming a fictitious inquiry with the type of question *material-finding* and both previous examples of general patterns assigned to *actors* [*particular group*] and *documentation* [*particular person*], then this inquiry would demand any kind of information object resulting from a documentation or observational activity conducted by a particular person targeting a particular group and its activities.

Moving on to the primary entities of interest identified during the interpretative analysis, in the case of general patterns from the provenance context, the possible primary entities of interest are any kind of actor since the scope of this study includes human-made traces and traditions. Consequently, and according to the systematization employed thus far, the kinds of

primary entities of interest identified are *particular person*, *particular group*, *type of person*, *type of group* and *actor*, in which case the creator or keeper is unknown or may vary.

In the case of a general pattern from the aboutness context the identified primary entities of interest are *particular person*, *particular group*, *type of person*, *type of group*, *particular activity*, *type of activity*, *particular thing* and *any*, which means that the general aboutness of the question or the specific aboutness of the information objects or facts is either unknown, irrelevant or variable.

Some factual and very few resource discovery inquiries have a topical relationship regarding aspects of particular information objects; for example, inquiries about the extent of a particular holding or the form or type of a document. The aboutness context is then given by the general pattern *Documents* and the primary entity of interest relates to a particular aspect of one or more information objects. In these cases, one of the following primary entities of interest is indicated, each of which subsumes various particular interests related to information objects:

- *content information* such as a summary of the contents of a file or holding, or the existence of a finding aid
- *document relation* such as whether a particular document is part of a particular file
- *extent* such as the number of documents within a file or holding
- *identity* such as the type or form of a document
- *provenance* such as the current custodian of a holding

The assignment of two general patterns, one describing the context of provenance and the other the context of aboutness, along with the specification of the respective primary entities of interest for each general pattern, allows for a categorization and approximation of the primary and most cognizable epistemic focus of an inquiry. A principal categorical query pattern can be described by instantiating elements from both general patterns with given and wanted entities from the inquiry and by denominating the primary query target.

The general patterns are not strictly delimited but overlap and constitute aggregations of entities and relationships with which to represent prototypical historical contexts related to provenance and aboutness. The general patterns and primary entities of interest can only provide an approximation of the most recognizable interest of an inquiry. The assignments do not constitute a clear-cut or exclusive categorization.

## 1.6 Fundamental Phenomena

As previously discussed (V:1.3), query patterns subsume several possible queries and not all possible indirections, or interpretations, but only the most relevant are either shown in diagrams or discussed in the accompanying text. Additionally, some basic modelling patterns of *fundamental phenomena* are omitted from the following discussions and diagrams in order to retain readability and reduce unnecessary complexity and redundancy. These fundamental phenomena

represent basic and reoccurring ontological structures which apply to all general patterns, such as the fact that activities may occur in sequence or that documents may form logical parts of other documents.

These fundamental phenomena are normally hidden, or implicit, in the diagrams and not explicitly discussed in the text since they typically pertain to any general pattern or query pattern. Such fundamental phenomena would need to be considered in query formulation in the context of an information system. In the context of this study, query patterns as exemplary instantiations of general patterns demonstrate the principle feasibility of representing interests of inquiries. As already mentioned, the purpose of the query patterns is not to constitute actual computational queries. Tzompanaki and Doerr (2012b) presented some fundamental patterns (by which they mean short-cuts) which hide more complex query patterns. Fundamental phenomena are similar and indeed, if proceeding to actual query implementation, such more complete representations would need to be considered. The focus of this study is, however, the theoretical and ontological level.

The *first* fundamental phenomenon pertains to actors; that is, *persons and groups* who carry out activities and actively or passively participate in events. If the inquiry clearly allows for the conclusion that either must have acted in the particular case, then this entity is used in the query pattern. However, in any case, the person or the group could have been a member of another group which may have carried out the activity or under whose authority the person or group may have acted. Therefore, as shown in Figure 22, a query will always have to consider a person and a group, both of which may be the member of one or more other groups, and all of which may have carried out an activity.

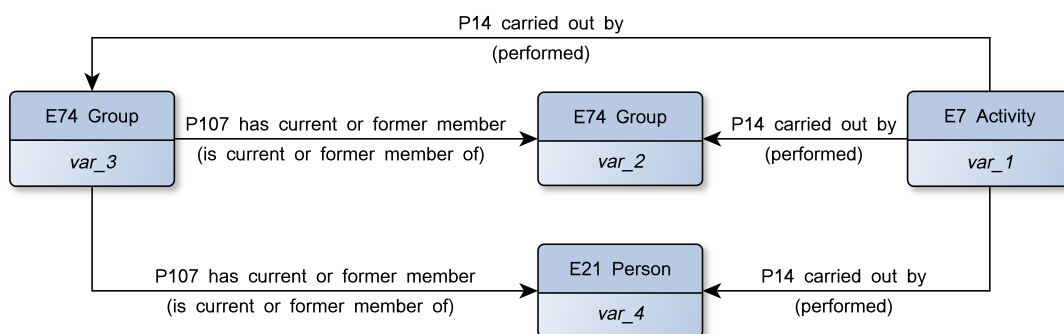


Figure 22 – Groups or persons carrying out an activity.

An inquiry may not allow for the conclusion as to whether a person or group must have carried an activity. In such cases, the query pattern will only contain an *E39 Actor*. Figure 23 shows the complete underlying pattern of the fundamental phenomena.

The class *E39 Actor* entails possible instances of the classes *E21 Person* and *E74 Group*, and, as previously discussed, membership in a group also needs to be considered. Figure 23 is thus a shorthand version of Figure 22.

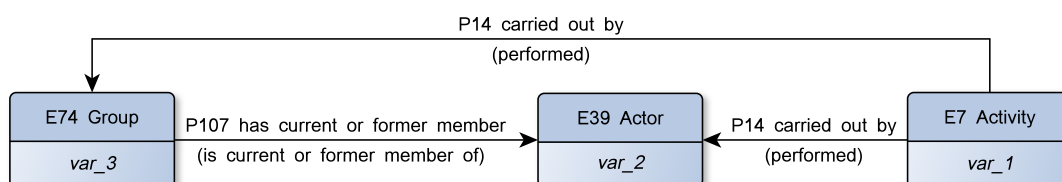


Figure 23 – Actors carrying out an activity.

The aforementioned fundamental phenomenon essentially also holds for the participation of persons and groups in events. Actors may do so as a single person, a group, or member of a group.

The *second* fundamental phenomenon pertains to *part-whole relations* and *sequences*. Events and activities as well as information objects and physical objects may be composed of sub-parts or they may be related to each other in some sequential order. Even though the sequences and part-whole relations are briefly introduced in the context of general patterns, these fundamental phenomena receive no further attention in the query patterns.

While the previous two fundamental phenomena are only hidden from the general and query patterns, the following are excluded from the discussion that follows, mostly due to their complexity, which would exceed the scope of this study and indeed be unnecessary from an illustrative point of view.

The *third* fundamental phenomenon pertains to *temporal* and *spatial entities*. As already mentioned during the introduction of the CRM (IV:3.2), the representation of time and place is not further explored. Time-spans during which activities or events occurred are simply represented using the class *E52 Time-Span* and the geographical location of an activity or event is represented using the class *E53 Place*. The fully fledged and adequate representation in both cases would require additional classes and properties, and entail further conceptual issues and use cases such as one time-span of an activity falling within another.

The *fourth* fundamental phenomenon pertains to the representation of particulars and types in the CRM. As already mentioned, the CRM allows for a form of type categorization that does not result in new classes but in a terminological differentiation of an existing instance. This facility can be used to address interest; that is, queries concerning particular things and kinds of the same thing. For example, the particular document with the minutes of a particular press conference versus all documents of the type “minutes” about all activities of the type “press conference”. In principle, this mechanism can be used everywhere and is not explicitly



represented in the general patterns. The class *E55 Type* is also used in this role in several specific cases already in the CRM, such as to represent the general use of an item.

Finally, the *fifth* fundamental phenomenon pertains to the representation of *appellations* and *identifications* of CRM entities. The instances represented in a knowledge base using the CRM can be named and provided with an identifier. The CRM provides the class *E41 Appellation*, which comprises “all proper names, words, phrases or codes, either meaningful or not, that are used or can be used to identify a specific instance of some class within a certain context” (Crofts et al., 2006, 20). Several sub-classes allow for the representation of specific kinds of appellations, such as actor appellation, place appellation, or time appellation, addresses, spatial coordinates, titles, or identifier. The general patterns neither explicitly represent such a level of detail nor will the query or archival patterns do so. This study refers to the facilities of the CRM for the representation of appellations and identifiers but will not discuss them further.

In particular, the identification and location of the physical items within the archive is explicitly excluded from the following discussion. The workings of archival identification systems for archival items, for example via call numbers, is neither discussed nor explicitly included in the general patterns.

The interest of the study at hand is not the identification of particular resources within an archival information or documentation system. The query patterns do not, therefore, make statements about the actual identification of physical items in an archive. For example, call numbers are only explicitly represented in the *Documents* pattern (V:2.2.6) in order to demonstrate that, in principle, all necessary archival identification information could be incorporated. Here too, the AKM does not include a detailed model of the archive. Its primary interest is in the “historical reality”, not the current “archival reality”.

The meaning and identity of instances of classes will be sufficiently apparent from their labels provided in the diagrams. In reality, such labels would have to be replaced by proper global identifiers and readable titles or texts that would allow for precise searches and string-matching. Their implementation is the task of cataloguing rules such as *Resource Description and Access* (RDA).<sup>67</sup>

The general patterns will therefore typically omit the aforementioned fundamental phenomena for better readability. The fundamental phenomena will normally be introduced only once at the first and most general level of occurrence.

## 1.7 Summary

The current section has reported on various results from the interpretative analysis, each of which constitutes important outcomes but also, at the same time, paves the way for the elaboration of the general patterns as the primary contribution of this study.

The initial, linguistically oriented analysis of the user and case files from the German

---

<sup>67</sup> <http://www.rda-jsc.org/>

Federal Archives and the National Archives of Norway resulted in 762 single inquiries. The determination of the *type of question* showed that 364 inquiries (48%), nearly half of the sample, were of the type *resource discovery* and 112 inquiries (15%) were of the type *fact-finding*. The remaining 286 inquiries (38%) were of the type *non-discovery*. Inquiries of the type *resource discovery* and *fact-finding* constitute the sample analyzed and interpreted further, in total 476 single inquiries (62%).

The next step in the analysis identified the referenced entities. The given entities are those referenced by propositions in the questions and the associated contextual information, and the wanted entities are those primarily and most recognizably referred to by the wanted archival or non-archival material, or by the wanted fact. They indicate the principal interest of the inquiry. In other words, the wanted entities are things about which the user is seeking factual information or information objects referring to these, and the given entities are those which contextualize the wanted entities and render them identifiable.

The primary categories of given entities identified were *Actors* with 609 occurrences, which equals 37% of all given entities, followed by *Time* and *Events* with 284 (17%) and 281 (17% of) occurrences respectively. *Documents* with 207 (13%) and *Places* with 168 (10% of) counts came next. *Other Entities*; that is, identifiers and general contexts as well as *Things* played only a minor role. The given entities have been further differentiated into particulars and types and will reappear as the basic building blocks of the general patterns.

The primary categories of the wanted entities were *Actors*, including persons and groups, amounting to 292 counts which equals 61% of all wanted entities. Next came *Events / Activities*, including unintentional events and activities, where unintentional events were near to insignificant with only 2 occurrences, compared to 120 activities which equal 25% of all wanted entities. *Documents* with 43 (9%), *Things* with 12 (2.7%), *General Topics* with 4 (0.8%), and *Places* with only 2 (0.4% of) occurrences were much less important.

The determination of the referenced entities made it possible to significantly reduce the complexity of the phenomena in the inquiries. Further, the wanted entities already indicate the perceived immediate ontological interest of the inquiry. In combination with the type of question they allow for an assessment of the general interest of the inquiries, where the type of question indicates whether the user is first and foremost seeking information objects or facts, and the wanted entity constitutes an approximation of what the inquiry is about in general as well as what the desired information objects and facts are about in particular.

However, the wanted entities constitute only a preliminary understanding of the users' interests and need to be interpreted further in order to arrive at a more thorough understanding and explicit representation of archival user needs. On the *ontological* level, the inquiries refer to or are *about* a specific historical context which may be expressed and described more or less implicitly or explicitly, in greater or lesser detail. The interpretative analysis of the inquiries renders this historical context explicit by formally describing its relevant entities in the AKM.

The interpretative and ontological analysis of the inquiries reveals three *levels of interest*:

*material fact*, *psychological* and *collective statistical*. These levels indicate how immediately the interest of an inquiry can be translated into an ontological model adequately representing the subject matter of their interest in the form of material facts. Most questions (399) are of the type *material fact* which amounts to 84% of all inquiries. Only 53 (11%) are collective statistical and 24 (5%) are psychological. On the one hand, most inquiries seem to contain many questions already fit for appropriate representation, i.e. as material facts, and on the other hand one can expect to be able to translate and answer most of these questions more or less directly.

The continuing analysis further results in an *ontological core framework* representing the most dominant entity types and relationships in the domain of discourse of the analyzed inquiries. As a point of reference for the interpretation of the inquiries, the framework guided the ontological modelling process. The general patterns are derived from the framework by further extending and describing segments of the historical reality described. Dominant entities were information objects as well as plans and wills, acts and actors.

While ontologically modelling the general patterns, two fundamental kinds of historical contexts emerged. The *provenance context* generally describes how things have been created or produced, who kept things, or to whom these things have been designated or sent. Within the scope of this study, these things are mostly textual information objects. The *aboutness context* generally describes the historical reality to which these information objects relate in a descriptive or topical sense. Both the provenance context and the aboutness context are principle ontological categories which subsume several concrete general patterns. While the AKM as such comprises a range of interrelated entities (entity types and properties as well as scope notes) a *general pattern* represents a particular aggregation of adequate entities from the AKM, describing typical contexts or situations, pertaining either to the provenance context or to the aboutness context. Each general pattern therefore belongs to one of the two contexts.

In the case of *resource-discovery questions* inquiring about any kind of information object (material-finding, research question, consultation), specific types of information object (specific type), or a particular information object (specific item), it would appear reasonable to describe the context of provenance or the topical aboutness of the desired information object. However, in the case of *fact-finding questions*, the fact may be found in either context. Since this study assumes a fundamental need for evidence for the requested factual information (IV:2.1), information objects potentially capable of providing such evidence are implicitly considered secondary objects of the interest of the question. Fact-finding questions are therefore always described with both contexts.

The next section will present the general patterns that have been identified based on the method previously discussed and the insights obtained thus far from the interpretative analysis of written natural language inquiries placed of archives. The discussion will begin with the general patterns belonging to the provenance context and then proceed to those from the aboutness context.

## 2 General Patterns

In this section, the general patterns will be introduced along with exemplifying query patterns.<sup>68</sup> The general patterns are the result of the deeper interpretative analysis of the inquiries and of the subject matter of their interests. The CRM has been used as a means to formalize and explicate these results in the form of ten general patterns, which together constitute an ontological model, the AKM. In the first sub-section, the general patterns from the provenance context will be discussed. The second sub-section will introduce the general patterns pertaining to the aboutness context.

Each general pattern will be introduced with an informal description of its scope and a diagram showing its ontological structure. Then each component of the general pattern will be discussed in more detail, followed by one or more query patterns providing additional and exemplary explanations.<sup>69</sup> Since query patterns always instantiate entities from two general patterns, one from the provenance context and one from the aboutness context, those query patterns that are provided first in the context of the *Provenance* patterns will necessarily anticipate the components used from general patterns from the aboutness context, introduced in detail later. However, the examples have been chosen in such a way to render them comprehensible without further explanation of the aboutness context. Further, as already mentioned, the query patterns are exemplary in that they typically indicate more than one potential query adequate to retrieve relevant information objects or facts. The examples are neither exhaustive nor comprehensive. The presentation of each general pattern closes with a statistic on its occurrence within the sample and a summary of its semantics.

### 2.1 Provenance Context

The general patterns that are part of the pattern group *provenance context* focus on the context of provenance of the items inquired about by a user. They address basic questions such as how, when or where things have been created, by whom, and where they have been kept. Table 4 provides a listing and informal scope notes of the four general patterns which describe the provenance context: provenance, documentation, self-documentation and correspondence. The dashes indicate a sub-pattern relationship in so far as documentation, self-documentation and correspondence describe more specific historical contexts than the generic *Provenance* pattern which constitutes the uppermost layer in the provenance context.

---

<sup>68</sup> For the notation used in the diagrams see the previous chapter (IV:3.3).

<sup>69</sup> The single inquiries have been assigned identifiers for easy reference throughout the text. The identifier starts with “Q” for “question” followed by three blocks of numbers: the first block holds a running number pertaining to a user or case file in the order of their processing; the second block holds a running number pertaining to single letters or emails within a user or case file; the third block then holds a running number pertaining to a single question within a single letter or email. The identifier is followed by the abbreviation “barch” or “nan” indicating its origin. For example: Q021-02-04<sub>barch</sub> indicates the fourth question within the second letter or email within the 21st user file analyzed in the sample from the German Federal Archives.

<i>Pattern Name</i>	<i>Informal Scope Note</i>	<i>Page</i>
Provenance	A generic, yet inclusive general pattern focusing on the things by and of an actor, i.e. the creation and production of things, and who has kept these things. Forms together with the <i>Aboutness</i> pattern the uppermost layer of the AKM.	131
– Documentation	Describes the documentation of events and activities carried out by persons and groups in the form of documents. Sub-pattern of the <i>Provenance</i> pattern.	143
— Self-Documentation	Describes the self-documentation of one’s own activities. Sub-pattern of the <i>Documentation</i> pattern.	154
– Correspondence	Describes the designating, addressing and taking notice of things. Sub-pattern of the <i>Provenance</i> pattern.	162

Table 4 – General patterns of the provenance context.

### 2.1.1 Provenance

All things stored in an archive have been created and kept by one or more persons or groups in the past. Statements about the creation and production as well as the possession of things are the most fundamental statements to be made about archival material.

The pattern *Provenance* forms, together with the pattern *Aboutness* (V:2.2.1), the uppermost layer of the AKM in terms of its semantics. All other patterns are specializations of these two general patterns.

**General Pattern** The pattern *Provenance* covers the generic *context of creation and possession* of things or, in other words, the things *by* and *of* an actor. Three principal and generic cases of creation and possession of things can be distinguished:

1. Identifiable but *immaterial* things (*E73 Information Object*) which have been *intellectually created* by an actor. Examples are the conceptual *contents* of any kind of *textual* item such as a letter, poem, newspaper article, minutes, or of any kind of *visual* item such as a painting, drawing or photograph.
2. *Persistent* and *tangible* things (*E24 Physical Man-Made Thing*) which have been *physically produced* by an actor. Examples are letters, poems, newspaper articles or minutes in the sense of a *physical carrier*, such as, most commonly, paper on which textual or visual contents are written or printed. Accordingly, this means a painting, drawing or photograph in the sense of visual contents cast on a physical carrier such as a canvas, piece of paper, or photographic paper. Other examples of physically produced things by an actor which do not necessarily carry textual or visual contents include objects such as a carved wooden figure, drinking cup, or knife.

3. *Persistent and tangible things (E24 Physical Man-Made Thing)* which have been *kept* by an actor at some point during the actor's lifetime. Examples include all things physically produced by the actor as mentioned above but also other items such as letters received, party membership cards, birth certificates, collections of newspaper articles, or personal estates in the archive.

Figure 24 shows the general *Provenance* pattern. The pattern fundamentally distinguishes between *intellectual content* and its *physical carrier*. The following discussion will focus first on the *immaterial things* – the intellectual content – as instances of the class *E73 Information Object*.

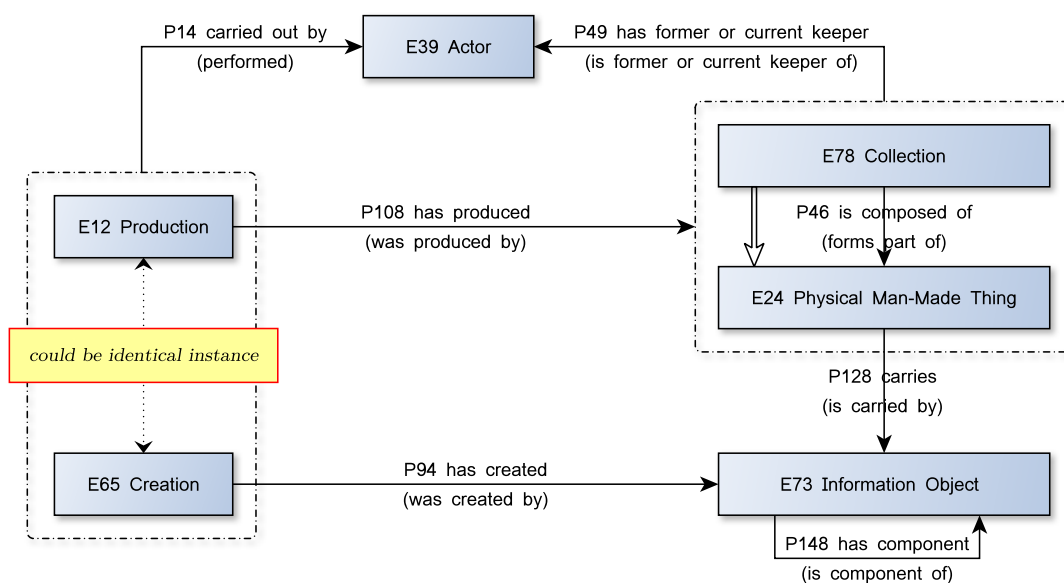


Figure 24 – General *Provenance* pattern.

Instances of the class *E73 Information Object* are identifiable but immaterial things in the sense of any kind of textual or visual contents. In other words, the class *E73 Information Object* comprises propositions in any symbolic form, such as words of a particular language, letters from an alphabet, or text characters.

These textual or visual contents can be further distinguished into instances of the classes *E31 Document*, *C3<sub>akm</sub> Expression of Will*, and *C4<sub>akm</sub> Plan* and its sub-class *C5<sub>akm</sub> Mandate*. As shown in Figure 25, these classes are sub-classes of *E73 Information Object*.

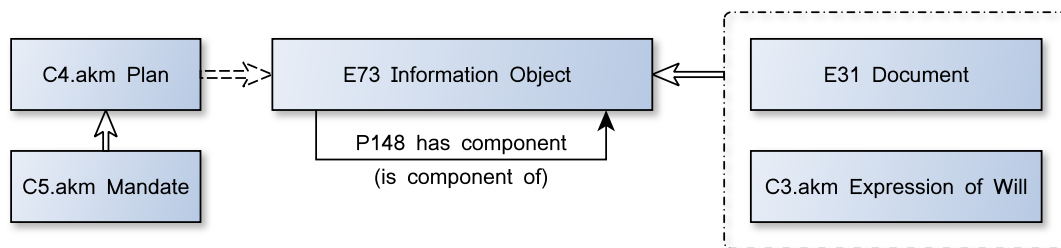


Figure 25 – Provenance pattern: Types of information objects created.

Each conveys specific notions of the nature of the immaterial contents they represent:

- Instances of *E31 Document* make propositions about reality which have a documentary character. These documents may be expressed in different forms, such as in a text, photograph, audio recording, diagram, etc. For example, the *minutes* of the parliamentary meetings of the Reichstag (1919-1930), the *surveillance reports* on revolutionary movements during the Weimar Republic, or the *photograph* of the handshake between Adolf Hitler und Paul von Hindenburg at the “Day of Potsdam” on March 21st, 1933.
- Instances of *C3<sub>akm</sub> Expression of Will* represent the identifiable but immaterial *will* of an actor to do something. For example, the vote (*C3<sub>akm</sub>*) of a parliamentary body to enact a law is the will to execute this plan. The class *C3<sub>akm</sub> Expression of Will* is a new sub-class of *E73 Information Object* and will be discussed in detail in the context of the *Plans* pattern (V:2.2.3).
- Instances of *C4<sub>akm</sub> Plan* are plans in the sense of schemas or descriptions for deliberate human activities or a product. Examples include manuals for operating devices, battle plans, city maps, social policies, architectural drawings, or laws. The class *C4<sub>akm</sub> Plan* has the sub-class *C5<sub>akm</sub> Mandate*, both of which are new classes as extensions of the CRM. Plans and mandates are discussed in greater detail in the context of the general patterns *Plans* (V:2.2.3) and *Documentation* (V:2.1.2).

The kinds of questions analyzed in this study do not inquire about specific textual or visual qualities of information objects. For example, no question asked for the language of a particular text or whether a photograph was in black and white or in colour, nor did any question require this kind of knowledge in order to satisfy a wider information need. In other words, the textual or visual nature of the contents of information objects did not entail specific relationships necessary to respond to particular queries.<sup>70</sup>

<sup>70</sup> Note again that classes are introduced only if they motivated specific relationships.

The classes *E33 Linguistic Object* and *E36 Visual Item* have therefore not been included. The former class comprises identifiable linguistic expressions which may be expressed in different natural languages such as texts in different languages, or the written or oral version of a text. The latter class encompasses visual items which convey information in the form of prototypical visual features in images, marks, or symbols such as a metro map as a diagram of lines and nodes, or particular derivative paintings of the original Mona Lisa painting.

Rather, this study concentrates on questions inquiring into relationships deriving from the fact that the contents are of a documentary nature, express wills, or have plan-related characteristics. The documentary nature or aboutness of contents does not depend on its textual or visual qualities and therefore does not need to be instantiated as an *E33 Linguistic Object* or *E36 Visual Item* but only as an *E31 Document*. The media type of an information object would be given by an instance of the class *E55 Type*. For example, a depiction of an inaugural ceremony would be an instance of the class *E31 Document* since it documents an activity, and the depiction could have the type “photograph” (*E55*) if indeed it were an example of photography.

None of the above information objects are to be mistaken for single items or the contents of a single page. Rather, they represent intellectual content which may stretch over several units of a physical carrier such as several pages. For example, an instance of the class *E31 Document* may represent any logical unit of an archival aggregation such as a file, series or record.

Instances of *E73 Information Object* may contain any other instance of *E73 Information Object*. This relationship is expressed by the property *P148 has component (is component of)*. For example, minutes (*E31*) may contain a plan (*C4<sub>akm</sub>*) decided upon during a meeting, or the resolution (*C3<sub>akm</sub>*) passed by vote by a parliamentary body to build new housing is documented in the official minutes (*E31*) of the parliamentary meeting during which the law was passed.

Since instances of the class *E73 Information Object* are immaterial they are always carried by an instance of the class *E24 Physical Man-Made Thing*.<sup>71</sup> This relationship is expressed by the property *P128 carries (is carried by)*. Instances of *E24 Physical Man-Made Thing* are persistent and physical things which have been purposefully produced by the activity of an actor. For example, the minutes of a parliamentary meeting are a physical document (*E24*) which carries (*P128*) the summary (*E31*) of that meeting. The photograph (*E24*) of the same parliamentary meeting carries (*P128*) a visual documentation (*E31*) of that meeting.

The instance of *E24 Physical Man-Made Thing* may carry more than one instance of *E73 Information Object*. For example, a newspaper (*E24*) typically carries many different single articles (*E31*). Accordingly, one instance of *E73 Information Object* may be carried by many different instances of *E24 Physical Man-Made Thing*; for example, the visual documentation (*E31*) of the parliamentary meeting is carried by the newspaper (*E24*) and the photograph (*E24*) taken during the meeting.

---

<sup>71</sup> Strictly speaking, instances of the class *E73 Information Object* may also exist independently from a physical carrier, for example as an oral tradition in the human memory. However, the study at hand focuses on tangible traces such as predominantly written records and to a lesser extent audiovisual or visual representations.



When, therefore, we speak of “the” minutes, “the” report, “the” photograph, or “the” letter, these things must be conceptualized in terms of the nature of their immaterial content (E73) and the nature of the physical carrier (E24) of this content. Users usually seem not to differentiate between these two qualities: when inquiring about a written document or a visual item, users typically tend to *think* of an instance of *E24 Physical Man-Made Thing*, the physical item, although their questions *refer* to an instance of *E73 Information Object*, the contents.

Indeed, the content (E73) is most relevant: Users tend first and foremost to be in search of information objects or, more specifically, documents, plans, expressions of will, or visual items, or any combination thereof, rather than specific physical carriers (E24) such as a photocopy or microfiche.

However, the differentiation between immaterial contents and physical carriers becomes relevant as soon as attributes of either are under scrutiny. For example, it makes a difference whether the historian obtains the original or a copy of a text which may have been transcribed by hand. The historian will always have to establish the intellectual and physical authenticity of the historical sources, which demands critique not only of a document’s contents but also of its form; that is, of the physical carrier. Authenticity will be discussed in the context of the *Documents* pattern (V:2.2.6).

In the *Provenance* pattern, persons and groups appear as actors creating or producing immaterial and physical things. The classes *E21 Person* and *E74 Group* are subclasses of the class *E39 Actor*. These concepts have already been introduced in the previous chapter (IV:3.2): The concept of actor (E39) comprises people, either as individuals or as groups, who “have the potential to perform intentional actions for which they can be held responsible” (Crofts et al., 2006, 19). Persons (E21) are individual people who have lived or are assumed to have lived, such as Karl Marx or St. Paul. Groups (E74) are any gatherings or organizations of two or more people, such as the German Communist Party (*Kommunistische Partei Deutschland*, KPD), or the Greens fraction in the Reichstag. Political offices such as the Chancellor of the Federal Republic of Germany (FRG), or the party leader of the Greens, as well as the concept of family are also considered a group.

Instances of the class *E73 Information Object* are created through instances of the class *E65 Creation*. This relationship is expressed by the property *P94 has created (was created by)*. Creation activities are always carried out by instances of the class *E39 Actor*. This relationship is expressed by the property *P14 carried out by (performed)*. For example, within the scope of this study, a common and generic case is the conception (E65) of the intellectual content (E73) of a text, such as the contents of a personal letter, by the author (E21) of that letter. In a more abstract sense, the contents (E73) of minutes are created (P94) by a group (E74) by talking (E65) with each other during a meeting.

Instances of the class *E24 Physical Man-Made Thing* are produced through instances of the class *E12 Production*. This relationship is expressed by the property *P108 has produced (was produced by)*. Production activities are always carried out by instances of the class *E39 Actor*.

This relationship is expressed by the property *P14 carried out by (performed)*. For example, the contents of a letter (E33) are written down (E12) on a piece of paper and thus the physical letter (E24) is produced. Similarly, a group, or a member of this group, writes down (E12) the summary (E31) of a meeting with a typewriter and thus produces the physical minutes (E24). The relationship between the two activities *E65 Creation* and *E12 Production*, therefore, is that the first conceives intellectual content (E73) which is put on a physical carrier (E24) by the second. However, in most cases, the instances of *E65 Creation* and *E12 Production* are identical. For example, the activity where a person writes a letter is an instance of *E65 Creation* and of *E12 Production* since the person conceives of the content and writes the actual letter at the same time, thus simultaneously creating and producing instances of *E73 Information Object* and *E24 Physical Man-Made Thing*.

Similarly, the participants of a meeting can be seen as collectively creating and producing the minutes of that meeting by talking (E65) with each other and then by writing (E12) the minutes down. This process, which is discussed in more detail in the *Self-Documentation* pattern (V:2.1.3), can be seen as a coherent activity, even if the minutes are actually written down later after the meeting has finished, and by only one member of the group. The person taking the minutes, of course, may be regarded as performing a distinct *E65 Creation* activity since this person writes down his or her own interpretative summary of what has been said during the meeting. Instances of one person dictating a text to another are also conceivable. However, although both cases can be represented, this level of complexity yields no added value since users are not explicitly interested in finding documents based on who conceived the intellectual content and who produced the physical carrier of this content.

As a side note, in FRBRoo, “a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information” (Bekiari et al., 2010, 10), the class *F28 Expression Creation* is a sub-class of both *E12 Production* and *E65 Creation* and thus unifies the processes of the physical and intellectual creation of things. This confirms the point made previously that it is generally not essential to distinguish between E12 and E65.

However, on the generic level of the *Provenance* pattern, the distinction between *E65 Creation* and *E12 Production* is useful and necessary for two reasons:

1. As already mentioned, not all instances of *E24 Physical Man-Made Thing* necessarily carry an identifiable instance of *E73 Information Object*; for example, objects such as a wooden walking stick or a tea cup may not carry an instance of *E73 Information Object* and result only from an instance of *E12 Production*. Even if it is rare that an inquiry specifically asks for physically made things, if all things *by* an actor should be described by the AKM and thus records that are theoretically retrievable, such objects need to be included.
2. More importantly, the act of *reproduction* is a special case of *E12 Production*. For example, a surveillance agency has collected copies of letters, or a law enforcement agency has collected evidence during an investigation which may include copies of archival materials.

Again, although no evidence was found during the analysis of the inquiries that users seek specific reproductions, i.e. physical carriers, of an information object, it is pivotal for historical and archival research to establish the physical and intellectual authenticity of the items requested, as has been discussed in the previous chapter (IV:2.1). For this purpose, documentation of reproduction activities and additional context information about the who, when and where of the reproduction are important. Since intellectual and physical authenticity pertain to the intellectual and physical features of a document, the pattern *Documents* (V:2.2.6) covers reproduction activities in more detail.

As shown in Figure 26, instances of the classes *E12 Production* and *E65 Creation* may take place at (*P7 took place at (witnessed)*) a geographical place (*E53 Place*) and happen (*P4 has time-span (is time-span of)*) during a specific time-span (*E52 Time-Span*). Furthermore, they may fall within a particular period (*E4 Period*). Time and place are fundamental historical categories and therefore must be represented in the model. While users in this study do not inquire about the time or place of creation or production activities as *wanted* information, they often provided these aspects as *given* information.

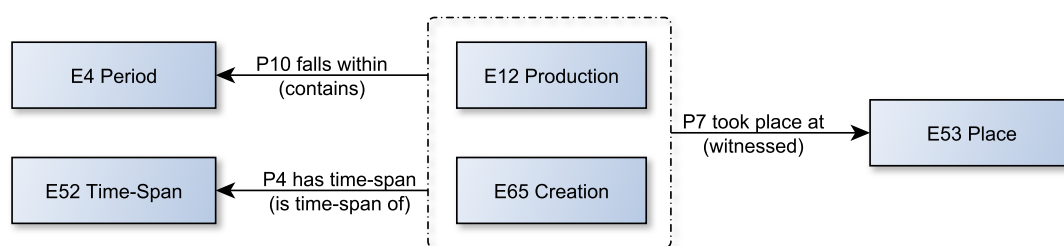


Figure 26 – Provenance pattern: Time and place for creation and production activities.

The pattern *Provenance* not only covers acts of creation and production but also the physical possession of things. The property *P49 has former or current keeper (is former or current keeper of)* covers cases where an instance of *E39 Actor* physically kept an instance of *E24 Physical Man-Made Thing*. For example, these could be things designated for the actor and received and then kept by him or her, such as a party membership card, a passport, or letters received from another actor.<sup>72</sup> One of the main reasons for including *E24*, therefore, was that questions about the possession of things require a query into instances of *E24 Physical Man-Made Thing* since only physical things can be physically kept or stored by an actor.

Instances of the class *E78 Collection* are curated and preserved aggregations of one or more instances of the class *E24 Physical Man-Made Thing*. This relationship is expressed by the property *P46 is composed of (forms part of)*. Similar to *E24 Physical Man-Made Thing*, an instance of *E78 Collection* is initially produced (*P108*) by an instance of *E12 Production* or may be kept

<sup>72</sup> The specific case of *designation* is covered in the pattern *Correspondence* (2.1.4) which is a specialization of the *Provenance* pattern.

(P49) by an actor. Note that *E78 Collection* is a sub-class of *E24 Physical Man-Made Thing* and therefore represents a material, and not a conceptual or logical, structure. For example, an administrative department (E74) created (P108) a collection (E78) of newspaper articles (E24) for the daily early-morning departmental briefing. In such cases it is reasonable to say that the administrative department was both producer and keeper of the collection, and the newspaper articles the collection contained. The department was not, however, the intellectual creator or physical producer of the newspaper articles. No assumptions are made about how an *E39 Actor* might have come into physical possession of an instance of *E24 Physical Man-Made Thing*. Furthermore, physical possession does not entail legal ownership of the things kept. The interest is not whether an actor legally owned an item or how he or she might have obtained custody of it, but only whether or not the actor physically kept the object at some point or other.

The *initial* creation of a collection is represented by an instance of *E12 Production*. In the context of the *Provenance* pattern the initial creation of a collection is not relevant. However, in the context of the *Documents* pattern (V:2.2.6), the initial creation of a collection is implicitly represented when the things of a person are transferred into the custody of the archive and aggregated into a *personal estate*, which is a special kind of collection. In the CRM, collections are curated by instances of *E87 Curation Activity* which may involve the addition (*E79 Part Addition*) and removal (*E80 Part Removal*) of things. Instances of *E6 Destruction* describe the destruction of a collection. The analysis of the questions showed that these aspects of curation introduce a level of complexity irrelevant to users' questions and therefore not represented in the model.

As the first example will show, questions about things *of* or *by* a person, specifically *personal materials* or *personal documents*, are a typical instantiation for the *Provenance* pattern. A query should result in the identification of all things which have been created, produced or otherwise kept or aggregated by a person during his or her life. In an archive, *personal estates* would specifically contain such *personal* things of or by an actor. Collections as personal estates would therefore be part of a query about the personal things of a specific person. Since personal estate is a quality given to a collection on account of its custodial acquisition by an archive, both patterns *Provenance* and *Documents* are closely related in this regard. In cases where a question inquires into the (personal) things of or by a person, the *Documents* pattern may be assumed as a third supplementary pattern, since it is important for covering the question.

#### **Example - "Any personal materials by a particular person"**

Context                    "*First wife of [person name 1] (born as [person name 2 (maiden name)], 1899 in HH, died 1990 in Berlin).*"

Q003-06-02<sub>barch</sub>    "*In which Berlin archive can I find personal materials regarding [person name 3]?*"

The question seeks archival material regarding a specific person. However, the general type of question is *resource discovery <consultation>* rather than *resource discovery <material-finding>*, since the knowledge of the archivist about archives is specifically addressed.

The *wanted* entity is *<person.regarding>*, more specifically personal materials of a specific person. The *given* entities include *particular person* (“[person name]”, “[person name (husband)]”, “[person name (maiden name)]”), and *type of document* (“personal materials”). The *supplementary* information includes *particular place* (“HH”, “Berlin”), *time-span* (“1899”, “1990”), and *type of event* (“death”).

The question is a *material fact* question since it inquires about the location of a certain type of document. The *provenance context* is *provenance [particular person]* since no further indication is given other than that the “personal materials” were either created, produced, or kept by a particular person.

The *aboutness context* is *aboutness [any]* (V:2.2.1) since no indication is given to which entity or topic the “personal materials” relate. The interpretation of “personal materials” by a particular person includes all the intellectual and physical things that have been created, produced, or kept by that person.

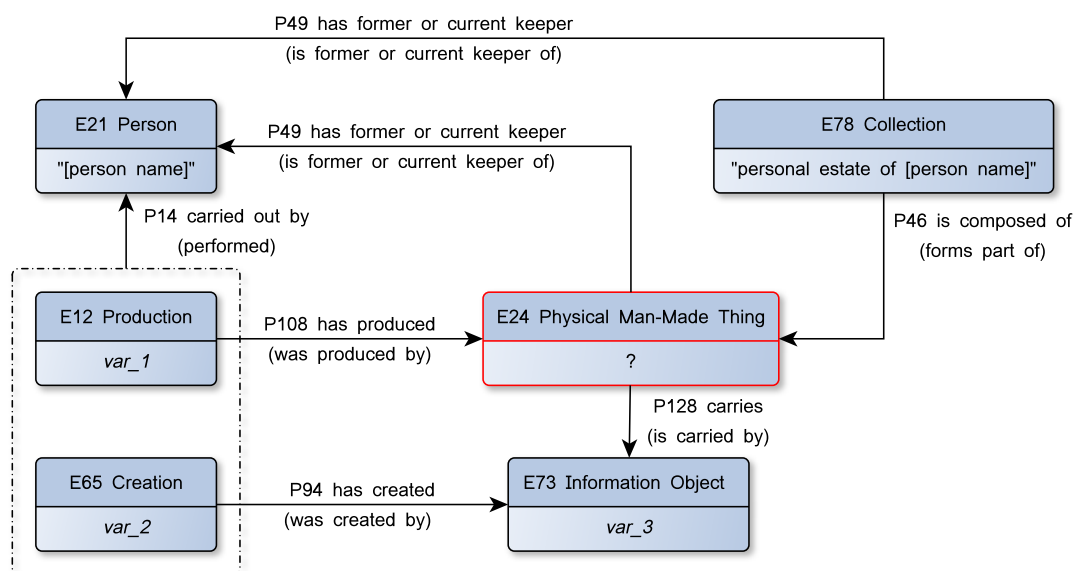


Figure 27 – Q003-06-02<sub>barch</sub>: Any personal materials by a particular person.

Figure 27 summarizes possible instantiations of the query pattern. Five sets of distinct queries can be distinguished which, together, will retrieve a relevant set of “personal materials” of the person in question.

1. The query asks for any physical thing as the result of an activity carried out by the person in question and which is, at the same time, an instance both of the class *E12 Production* and of the class *E65 Creation*. As previously discussed, since the person conceived of the intellectual content and simultaneously put this content onto a physical carrier, these

things can be considered *original things*. Examples include personal diaries, personal letters to friends, or letters to a municipal or governmental authority.

2. The query asks for any physical thing produced by the person; for example, a self-made walking stick or watercolour paintings.
3. The query asks for any physical thing which carries the intellectual content created by the person, even when the physical carrier has not been produced by that person. This case covers any kinds of reproduction or copies of the intellectual work of the person in question made by others. This means that the instances of the class *E12 Production* and *E65 Creation* are not the same and have been carried out by different actors. Examples include copies made of intercepted letters by a secret agency.
4. The query asks for any *single* physical thing which has been *kept* by the person in question but has not been intellectually created or physically produced by that person. Examples include party membership cards, passports, a birth certificate, medals, diploma, but also commodities such as glasses, a wallet, or a walking stick, if this walking stick has been purchased.
5. The query asks for any aggregation or *collection* of physical things which have been kept by the person in question. This includes, as a specific case, any *personal estate* as a archival unit in an archive. Strictly speaking, a query regarding a personal estate is part of the *Documents* pattern (V:2.2.6), since the attribute “personal estate” is a quality assigned to an aggregation of things during their accession into an archive.

No assumption is made about the legal ownership of things the person kept during his or her lifetime or about the specific type or form of the personal things created or kept. For reasons of simplicity, the additional information provided in the original question about the person in question is not represented in Figure 27. However, this information may help to identify this particular person. The query pattern could be easily adjusted to limit the search to a more specific, yet general type of “personal material” such as any thing (E24) carrying textual contents (E33) or any thing (E24) carrying visual content (E38) incorporated in a collection (E78) such as a “personal estate” (E55).

Finally, even though the question is intuitively targeted at the name of an archive, the actual interest of the user is to find archival materials. After relevant instances of the class *E24 Physical Man-Made Thing* have been identified, the query could proceed to retrieve the actual location of these items in an archive and the address of this archive. These cases are discussed in the *Documents* pattern (V:2.2.6).

### Example - “Any documents kept by a particular institution”

Context “The focus of my private research is the history of philately and the Philatelic Collectors’ Organisation 1933-45.”

Q059-01-04<sub>barch</sub> “I am searching for documents of the postal museum (Berlin, Leipziger Straße).”

The general type of this question is *resource discovery* <material-finding>. The *wanted* entity is <group.by>, i.e. any documents of the postal museum. The relevant *given* entities are *particular group* (“postal museum”). *Supplementary* information is *particular place* (“Berlin”, “Leipziger Straße”).

The question is a *material fact* question since it inquires about an observable relationship of creation and possession between an actor and documents. The *provenance context* is *provenance [particular group]* since the question asks for any documents of the postal museum. The *aboutness context* remains unspecified and therefore *aboutness [any]* (V:2.2.1) is assigned.

Figure 28 shows a possible query instantiation for this very general and unspecific question.

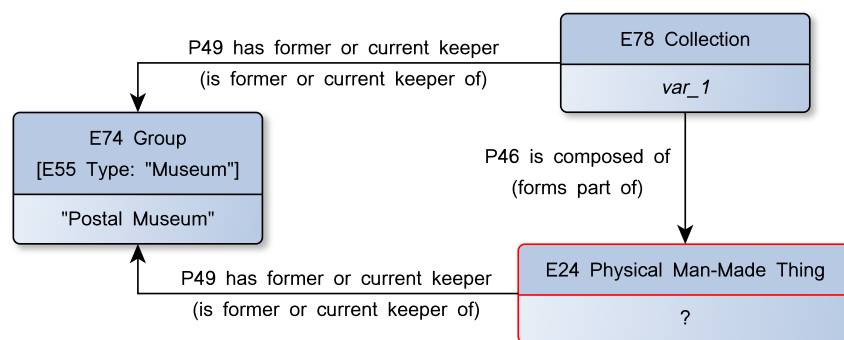


Figure 28 – Q059-01-04<sub>barch</sub>: All things kept by a group.

It may be reasonably assumed that the user is not specifically interested in things either created or produced by the “Postal Museum” but only in the things which have been kept by the museum. The query pattern then asks for two things:

1. any instance of the class *E24 Physical Man-Made Thing* kept by the *E74 Group* “Postal Museum”,
2. or any instance of the class *E78 Collection* kept by the *E74 Group* “Postal Museum”.

The relationship *P46 is composed of (forms part of)* is optional; however, it allows us to infer that if an instance of *E24 Physical Man-Made Thing* is found which has been kept by the “Postal Museum” and this instance is a part of a collection, then this collection can be regarded as potentially interesting even if it is unknown whether the collection itself has been kept by the “Postal Museum”. The supplementary information “Berlin, Leipziger Straße” has been omitted from the figure but it may help to identify the particular “Postal Museum”.

**Statistics** In the complete sample, the *Provenance* pattern has been assigned to 226 questions representing 47% of all inquiries. Of these 226 questions, 156 belong to the BArch sample representing 45% of all inquiries in that sample. The remaining 70 questions are from the NAN sample, representing 53% of all inquiries in that sample. The by far most frequent primary entity of interest is “actor”, with 171 occurrences in the whole sample, constituting 76% of all questions assigned to the *Provenance* pattern. In the BArch sample, “actor” appears 108 times (69%) with Provenance counts and in the NAN sample 63 times (90%). The second most important entity of interest is “particular group” with 38 counts (17%) in the whole sample, of which most belong to the BArch sample with 33 occurrences (21%) and only five in the NAN sample (7.4%). Particular persons are even less frequent with 11 counts (4.9%) in the whole sample, of which 9 (5.8%) appear in the BArch sample and two (2.9%) in the NAN sample. Types are near to irrelevant: “type of persons” does not appear while “types of groups” occur six times (3.8%) in the BArch sample.

<i>Provenance Context</i>	<i>All (n=467)</i>		<i>BArch (n=345)</i>		<i>NAN (n=131)</i>	
<b>Provenance</b>	<b>226</b>	<b>47%</b>	156	45%	70	53%
- Provenance [ actor ]	171	76%	108	69%	63	90%
- Provenance [ particular person ]	11	4.9%	9	5.8%	2	2.9%
- Provenance [ type of person ]	0	-	0	-	0	-
- Provenance [ particular group ]	38	17%	33	21%	5	7.1%
- Provenance [ type of group ]	6	2.7%	6	3.8%	0	-

Table 5 – Occurrences of the general pattern *Provenance*. Note that percentages for the primary entities of interest are per current general pattern.

The high number of actors as primary entities of interest in the context of the *Provenance* pattern is an indication that, in many cases, users either do not know specifics about the historical context – who created, produced, or kept an information object, for example – or that they deliberately aim for a larger number of search results by omitting specific information.

**Summary** The general *Provenance* pattern covers three generic notions of provenance: (1) The things *intellectually created* by an actor, (2) the things *physically produced* by an actor, both of which typically have identical instances, and (3) the things *kept* by an actor.

The pattern allows for the formulation of broad and inclusive queries for retrieving documents with regard to the creation or possession of an information object by an actor. The instantiation of all three notions covers questions for *any* things *of* or *by* an actor.

The pattern *Provenance* is assigned as a category to a question when the context of creation or possession is either not specified, which either means that no reasonable assumption can be made about the historical context of provenance to which the question is referring, or the



context of provenance is not relevant for the research interest of the question; for example, when questions inquire about documents about a specific topic.

As far as the aboutness of the documents requested is concerned, the general aboutness of questions and the particular aboutness of information objects is covered by the general patterns belonging to the *aboutness context* group.

The general pattern *Provenance* describes the simple context of creation and production of intellectual and physical things and the plain fact of possession of physical things by actors. The pattern does not further qualify these contexts other than by allowing for the specification of the temporal and geographical extent of these activities. The context of creation is more specifically described by three sub-patterns: *Documentation* (V:2.1.2), *Self-Documentation* (V:2.1.3), and *Correspondence* (V:2.1.4), each further detailing how a document has been created.

### 2.1.2 Documentation

While the *Provenance* pattern remains quite unspecific, though inclusive, in terms of the actual context of creation and production of things, the *Documentation* pattern specifies *how* things; that is, the intellectual contents of things, have been created. The pattern describes a fundamental activity where actors document and observe the activities of other actors and produce information objects about these activities. In fact, such documentation activities, which are typically based on official mandates, create the vast amount of archival materials in an archive.

**General Pattern** The *Documentation* pattern covers the *context of creation* of documents as the result of direct or indirect observation and documentation of events or the activities of actors. Three principal senses can be distinguished:

1. The documents (E31) created by direct or indirect documentation or observation of an unintentional *event* (E5). For example, the death (E5) of a person is documented by issuing a death certificate (E31).
2. The documents (E31) created by direct or indirect documentation or observation carried out by an actor documenting the *activities* (E7) of another actor. For example, a secret agency observes the political activities (E7) of a person through surveillance, and generates a report (E31) or newspaper article (E31) describing the events of a previous bombing raid (E7).
3. The *mandate* (C5<sub>akm</sub>) on which a direct or indirect documentation or observation may be based. For example, a parliament (E74) documents its own meetings based on an official mandate, for instance according to the respective parliamentary law (C5<sub>akm</sub>). On the other hand, a person will document his or her own life in a diary without any kind of explicit mandate.

Figure 29 shows the *Documentation* pattern. The core of the pattern is constituted by the classes C1<sub>akm</sub> *Documentation* and C5<sub>akm</sub> *Mandate*. Both classes are extensions of the CRM.

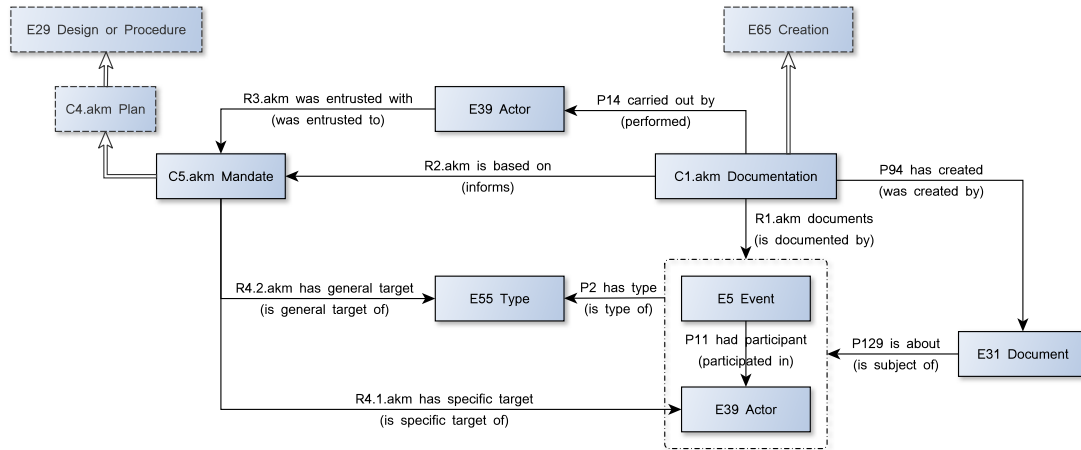


Figure 29 – General Documentation pattern.

The class  $C1_{akm}$  Documentation constitutes the heart of the pattern. Instances of  $C1_{akm}$  Documentation observe or document, in the most general sense, instances of  $E5$  Event or, in most cases, instances of  $E7$  Activity.<sup>73</sup> In this regard, the class  $C1_{akm}$  Documentation is very similar to the class  $F26$  Recording in FRBRoo which is “intended to convey (and preserve) the content of one or more events” (Bekiari et al., 2010, 50). The relationship between such documentation activities and an event or activity is expressed by the new property  $R1_{akm}$  documents (is documented by).

Examples of documentation activities include the documentation of the birth of a human being by issuing a birth certificate, the creation of a dossier about a past group of people, or the act of surveillance of the activities of a person. The former example describes the rare but relevant case of a unintentional event, the birth of a human being. As the analysis of the referenced wanted entities indicates, users do not directly inquire about unintentional events such as plagues, accidents or natural disasters.

However, the given entities showed that users indirectly provided many instances of a type of event, namely birth and natural death events. Such events are typically documented, for example, by birth or death certificates, both literally as a document of the birth event or death event of a person. It is therefore important that the model allow for the representation of unintentional events (V:2.2.2). In most cases, however, documentation activities provide a record of deliberate human activities as instances of the class  $E7$  Activity.

The two latter examples indicate that documentation activities may not only be direct observations but also indirect and a posteriori in relation to the documented event or activity. The documentation activity may occur long after the documented event or activity has taken place. Furthermore, documentation activities may subsume a series of direct and indirect acts of observing and documenting spread out over a period of time but which appear on the whole as

<sup>73</sup> The context of both events and activities is discussed in more detail in the *Events* pattern (V:2.2.2).

a coherent activity. The definition of what forms a coherent documentation activity in terms of geography, time, participation and objective, is a question of reasonable abstraction. Acts of surveillance, for example, may be modelled in detail, specifying many distinct documentation activities that are seen as appearing in a sequence and direct continuation of each other. The model allows for different levels of granularity, as described in the *Events* pattern (V:2.2.2).

Documentation activities are carried out by individual persons or groups; that is, by instances of *E39 Actor*. This relationship is expressed by the property *P14 is carried out by (performed)*. The activities (E7) that are documented or observed are carried out by different actors. The decisive criterion is therefore that one actor documents the active or passive involvement of another in an event or activity. For example, a surveillance agency (E74) conducts surveillance ( $C1_{akm}$ ) and documents ( $R1_{akm}$ ) the everyday activities (E7) of a specific person (E21), or a person (E21) has written ( $C1_{akm}$ ) a diary over several years documenting ( $R1_{akm}$ ) the events (E5/E7) witnessed.

Documentation activities typically result in instances of *E31 Document* which is expressed by the property *P94 has created (was created by)*. The things that are created during a documentation activity contain an account of the documented events or activities and are therefore instances of *E31 Document*. The relationship between a document and an event or activity is expressed by the property *P129 is about (is subject of)*.

As already discussed in the *Provenance* pattern (V:2.1.1), the creation (E65) of information objects (E73) takes place at (P7) geographical places (E53), happens (P4) during a time-span (E52), and may fall into (P10) a period (E4). These properties also pertain to  $C1_{akm}$  *Documentation* since this class is a sub-class of *E65 Creation*.

The pattern does not differentiate between instances of *E65 Creation* and *E12 Production*, as discussed and exemplified in the *Provenance* pattern (V:2.1.1), since the user is not interested in specific physical carriers (E24), for instance, the original or a reproduction on microfiche, but only in the document (E31) on an event or activity and its contents. Questions about originals and reproduction are covered by the *Documents* patterns (V:2.2.6).

The class  $C1_{akm}$  *Documentation* has been introduced as a new sub-class of *E65 Creation* because of the necessary documentary relationship between the documenting activity and the documented event or activity, expressed by the property *R1<sub>akm</sub> documents (is documented by)*. Since documentation activities typically create information objects, the class  $C1_{akm}$  *Documentation* is a sub-class of *E65 Creation* and not of *E7 Activity*. This is because only instances of the former class “result in the creation of conceptual items or immaterial products” (Crofts et al., 2006, 5) through *P94 has created (was created by)*. Finally, documentation activities typically follow a mandate, which is a specific kind of plan. All of these specific extensions to the scope of the class *E65 Creation* justify the introduction of the new class  $C1_{akm}$  *Documentation*.

Documentation activities are typically based on or follow mandates, formal codes of responsibilities, or given orders, all of which are instances of  $C5_{akm}$  *Mandate*. This relationship is expressed by the new property *R2<sub>akm</sub> is based on (informs)*. For example, the police conduct surveillance on a specific group based on a court order which legalizes the surveillance and

defines what is allowed and what not. Other examples of mandates include parliamentary rules of procedure, internal regulations of companies, or any standing orders.

Mandates formulate the principle scope of application for documentation activities; for example, they specify the targets of a documentation activity such as a particular actor, or a principal type of actor, event or activity. The target relationship towards a particular actor is expressed by the new property  $R4.1_{akm}$  *has specific target (is specific target of)*, while the target relationship towards a type of actor, event or activity is expressed by the new property  $R4.2_{akm}$  *has general target (is general target of)*.

Furthermore, the actor carrying out a documentation activity may have been entrusted with the mandate on which this documentation activity is based. This relationship is expressed by the new property  $R3_{akm}$  *was entrusted with (was entrusted to)*.

The class  $C5_{akm}$  *Mandate* is a sub-class of  $C4_{akm}$  *Plan* (V:2.2.3) which, in turn, is a sub-class of  $E29$  *Design or Procedure*. In general, instances of  $C4_{akm}$  *Plan* provide the structure of a plan as instructions for carrying out an activity. Mandates, however, are more specific insofar as they additionally and explicitly specify targets ( $R4.1_{akm}$ ,  $R4.2_{akm}$ ) for the execution of a plan and are entrusted ( $R3_{akm}$ ) to specific actors. They specifically comprise orders for the execution of documentation or observation activities such as, and most importantly in the context of this study, parliamentary rules of procedure which specify, for instance, that minutes are to be taken during parliamentary committee meetings. In contrast to instances of  $C4_{akm}$  *Plan*, mandates are explicitly authoritative and transpose an intention as well as authorization to conduct an act, typically in a regular and routine manner.

Mandates are one of the most crucial aspects for discovering and identifying potentially relevant documents since they provide indications as to which person or group might have had a reason (an official mandate) to write about a particular actor or to document certain types of activities. For example, two important general questions while searching for documents about a person in an archive are, as discussed in the previous chapter (IV:2.2): What were the important stages in the life of this person? And into the focus of which (public) entity or entities did the person come during his or her life?

This *mandate-based documentation (auftragsgemäße Dokumentation)* allows us to draw conclusions on the probability that specific types of events or activities have been documented and that documents can be expected in the archive. If an activity or event is known to have been documented by an instance of  $C1_{akm}$  *Documentation*, the existence of a document (E31) can be inferred with high probability. Consequently, the expectation to find a physical carrier (E24) of this document, either the original, a copy or a reproduction, in an archive is plausible, and increases if the documentation activity was based on an official mandate.

Furthermore, the mandate not only helps to discover documents but also allows for a determination of whether a document – as the result of a documentation activity – is of *official* or *non-official* character. Documents are official if they are the result of a mandate-based documentation activity.

The documentation activity and its contextual classes can be seen as part of a description of the *historical reality* as given in the user's question. The mandate, on the other hand, as a plan, belongs to an *intentional level (Absichtsebene)* as described in the context of the "Ontological Core Framework" (V:1.4). Here, principles are formulated which are meant to formally govern the historical reality and which might find their expression in documents. These documents are the point where this ontological representation of the historical reality would intersect with that of the *archival domain* of record keeping.

The documentation activity is closely related to the structure of an archive itself. The archive is essentially the physical result of many different kinds of documentation activities, including acts of self-documentation (V:2.1.3). Simply put, documents which have resulted from past (documentation) activities are preserved, organized in records, and described in archival aids by the archive.

These archival aids describe the structure of the collection and the (documentation) activities which have led to the creation of the documents, now part of records in the archive. In theory, the documentation activity pattern renders this context explicit and relates and contextualizes them with the historical context from which they have emerged.

Before introducing particular examples, a generic example should illustrate how the *Documentation* pattern works. Taking the case of surveillance, the typical representation would probably be of one documentation activity which is said to have documented any kind of activity of some particular actor or type of actor and which points to (has created) several sets of documents. These are files which contain, for example, reports. This documentation activity may be based on one or more mandates and may have been carried out by one or more actors (persons or, more likely, groups).

Here, and this is instructive, the documentation activity serves as a hub for connecting those who have carried out the documentation and written about the principal mandates, the actors and activities which have been documented and, most importantly, the documents which have resulted from this constellation. Any parts of such a pattern may be queried in order to retrieve relevant documents for questions which are concerned with retrieving relevant documents or facts, as the result of any kind of documentation activity.

### **Example - "Conducting surveillance on a type of group"**

Context                    "*One source I would like to consult are the police and surveillance reports for the Weimar Republic about revolutionary movements. I would like to know what the surveillance agency of the Reich had to say (...).*"

Q002-03-01<sub>barch</sub>    "*Do you know if the Bundesarchiv holds such documents?*"

The type of question is *resource discovery <specific type>*. The *wanted* entity is *<group.regarding>*. The *given* entities are *type of group* ("revolutionary movements", "surveillance agency of the

Reich”), *type of document* (“police reports”, “surveillance reports”), and *particular period* (“Weimar Republic”).

The question is a *material fact* question since the research interest aims at immediately observable and documented relationships. The provenance context is *documentation [type of group]* since the reports are the routine and official products of a state institution as the result of an observation of the activities of a group. The aboutness context is *actors [type of group]* since the reports are about a type of group and their activities.

Figure 30 shows a query pattern demonstrating how the question can be adequately represented as a query using the *Documentation* pattern.

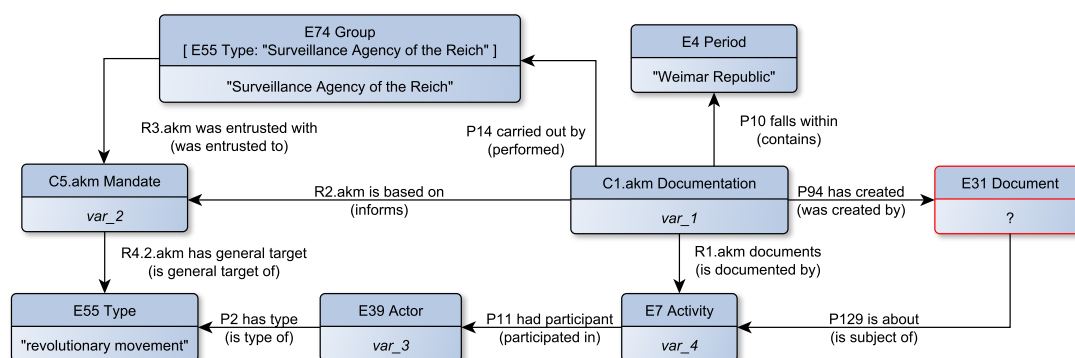


Figure 30 – Q002-03-01<sub>barch</sub>: Surveillance as an example of the observation of others.

The interpretation of the question is evident and materialized by the documentation activity (C1<sub>akm</sub>) in the centre of the figure. During the Weimar Republic (E4), a series of documentation activities were carried by a surveillance agency of the Reich (E74) which had been entrusted with a mandate (C5<sub>akm</sub>) to do so. The mandate had as its principle target an actor (E39) of the type “revolutionary movements” (E55). This actor conducted activities which were observed by documentation activities and documented in reports (E31).

The class *E39 Actor* is used here since it is unknown whether particular persons or groups participated in these activities. Furthermore, the property *P11 had participant (participated in)* is used in order to remain inclusive in terms of active or passive participation (P11) in activities (E7).

Keeping in mind that this is a simplified representation, this pattern expresses the formal basic structure of an answer adequate to satisfy the interest of the initial question. Documents resulting from this constellation are relevant documents and may adequately answer the user’s question.

Of course, the user actually wants to know if such documents are available in the German Federal Archives. The essential request implicit in nearly all resource discovery questions submitted to archives is for pointers to documents and the actual location of these in the archive, such as a set of call numbers. However, the AKM does not address the issue of identification of

documents for retrieving the actual physical document but rather describes the adequate context of documents and facts of interest for documentation and retrieval.

### Example - “Keeping record on a type of group”

Context                    “I am looking for files regarding Swedish citizens who were employed as lecturers in Swedish at a German university between 1933 (or earlier) and 1945.”

Q050-01-04<sub>barch</sub> “Did the Gestapo or similar organizations keep files on foreign citizens?”

The contextual information provides an implicit request for files regarding Swedish citizens employed to teach Swedish at German universities. The explicit question on the Gestapo will, however, be used as an example here.

The general type of question is *fact-finding* since the user wants to know whether certain organizations kept files on foreign citizens. The *wanted* entity is *<group.activity>*, i.e. general information about a mandate of the Gestapo or “similar” organizations to document the activities of foreign citizens. The *given* entities for the question are *particular group* (“Gestapo”), *type of activity* (“keeping files”), and *type of group* (“foreign citizens”).

The question is a *material fact* question since it inquires into the existence of files which are about a specific type of group. The provenance context is *documentation [particular group]* since the files in question are the result of a mandate-based observation of the activities of foreign citizens. The aboutness context is *plans [type of activity]* since the question revolves around the fact that some actor had a mandate to perform some type of activity.

Figure 31 shows an adequate query pattern for the research interest of the question. As in the case of every fact-finding question, the query asks for facts, and documents as evidence of these facts. Two principal queries can be distinguished:

1. The existence of a mandate (C5<sub>akm</sub>) which has been entrusted to the actor “Gestapo” and which targets any group of the type “foreign citizens”: Here, either the material fact may suffice that such a mandate exists in the knowledge base, or, additionally, the particular information object would be the relevant item for retrieval.
2. Documents resulting from a documentation activity carried out by the Gestapo and which have documented the activities of groups of the type “foreign citizens”: This would provide indirect evidence that the Gestapo did indeed have a mandate to “keep files” on “foreign citizens”.

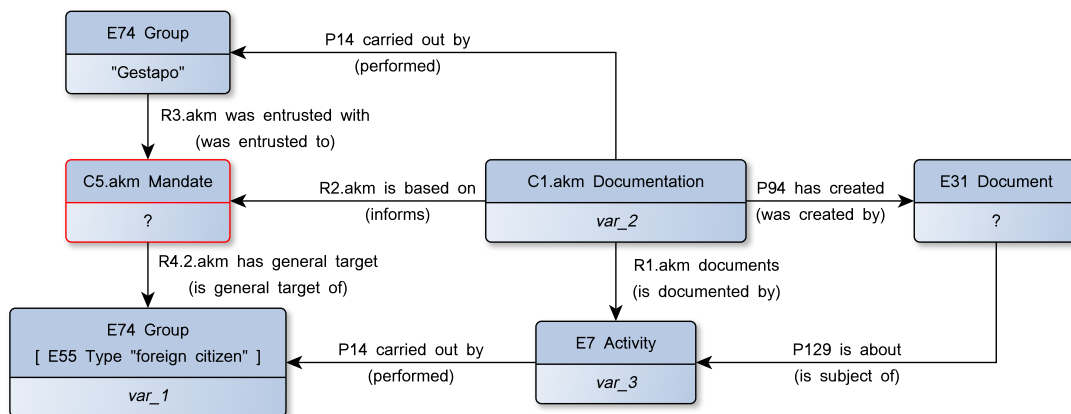


Figure 31 – Q050-01-04<sub>barch</sub>: Keeping records on a type of group.

The class  $C1_{akm}$  Documentation in the centre of the figure represents the hub which connects all other relevant contextual information. The particular documentation activities are unknown and insignificant for the retrieval of instances of the class  $C5_{akm}$  Mandate which fulfil the condition that they are targeted at groups of the type “foreign citizen”, have been entrusted to the particular group “Gestapo”, and have been used in documentation activities ( $C1_{akm}$ ) recording the activities of foreign citizens.

The main target of the question is the mandate ( $C5_{akm}$ ) meeting the particular conditions formulated in this query pattern. As previously discussed, in cases of fact-finding questions the need is assumed to find documents providing further evidence for the fact.

The secondary target of the query pattern is therefore documents (E31) resulting from a mandate for a documentation activity based on the mandate in question. This means that, even if the mandate is unknown, documents that meet the context formulated in this query pattern would indicate that the Gestapo indeed had the mandate to keep files on “foreign citizens”.

Another possible third query is for “similar organizations”. Here, background knowledge from controlled vocabularies or other knowledge bases would allow for an easy adaptation of the query to “similar organizations”.

### Example - “Creating a dossier on a type of group”

Q011-01-04<sub>barch</sub> “Furthermore, the SED had a file on anti-fascist activists during the time of National Socialism, which I would like to see as well.”

The general type of question is *resource discovery* <specific item>. The *wanted* entity is <group.regarding>. The relevant *given* entities in this question are *particular group* (“SED”), *type of group* (“antifascist activists”), *particular period* (“National Socialism”), and *type of document* (“file”).

The question is a *material fact* question since it inquires about the observable relationship between a document and its topic. The *provenance context* is *documentation* [*particular group*] since



a particular group, here the SED, created a dossier on another group through the documentation of their activities. The *aboutness context* is *actors [type of group]* since the file in question is about a type of group.

The principal interpretation of this question is very similar to that of the questions on pages 148 and 150, insofar as an actor, here the group or a member of the group SED, documented the activities of another actor, here persons of the type “anti-fascist activist”, during the period of National Socialism by creating a file on their activities.

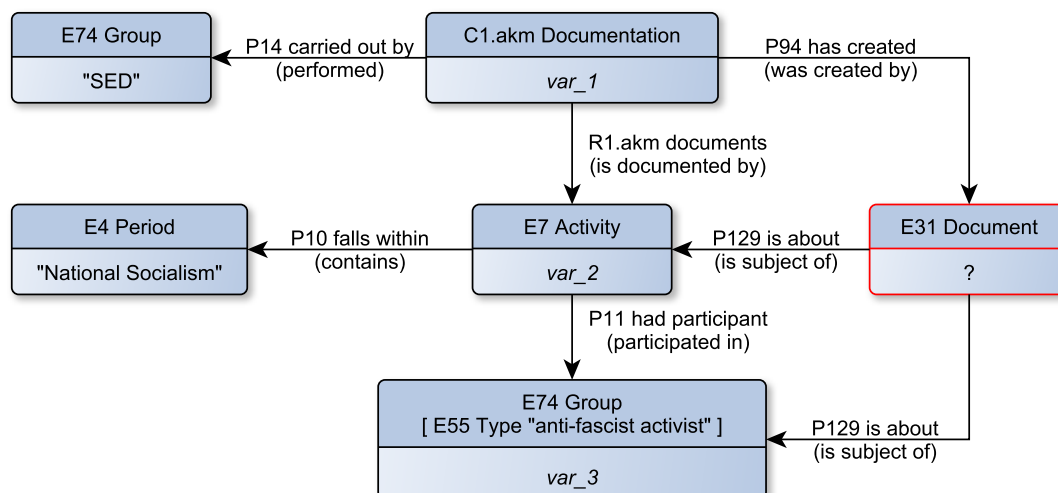


Figure 32 – Q011-01-04<sub>barch</sub>: Creating a dossier about anti-fascist activists.

The nature of the documentation activity shown in Figure 32, however, is different. While in the questions on pages 148 and 150 the documentation activity happened more or less at the same time or in parallel to the documented activity; in this example, the file, or better the contents of the file represented by *E31 Document*, was probably created not only during the National Socialist period but its supplementation continued even after the end of this specific time-span. In this regard, it is an indirect and probably a-posteriori documentation.

This is an example demonstrating that the documentation activity and the documented activity do not have to occur synchronously. The documentation activity may even start after the end of the documented activity. In such cases, as in this example, the documentation is indirect and resembles the collection of traces of the activity in question.

The query pattern assumes that the documentation activity focused on the past activities of the anti-fascist activists and not on personal features. In principle, a documentation may also record ( $R1_{akm}$ ) the personal features of actors, such as their hair colour, age or profession. In such a case, the property  $R1_{akm}$  documents (is documented by) would also point to the actor (E39).

**Example - “Documenting a particular activity”**

Q054-04-01<sub>barch</sub> “I am looking for documents regarding the denazification of [person name] (1871-1949).”

The general question type is *resource discovery <material-finding>*. The wanted entity is <activity>. The relevant given entities are *particular activity* (“denazification of [person name]”), *particular person* (“[person name]”), *type of event* (“birth”, “death”) and *time-span* (“1871-1949”).

This is a *material fact* question since it inquires about an observable topical relationship between a document and an activity. The *provenance context* is *documentation [actor]* since the documents in question must have been the result of some official documentation of the denazification of the person. However, since we do not know who actually conducted the documentation, we assume the identity of the unknown actor for the *Documentation* pattern. The *aboutness context* is *events [particular activity]* since the person is the subject of the denazification activity and not an active agent.

The interpretation of this question brings us to the conclusion that any relevant documents have been the result of a documentation activity that recorded the specific activity of the denazification of “[person name]”. Since such activities were official acts, the documentation must have been based on a mandate principally targeted at activities of the type “denazification” and, of course, the particular person “[person name]” must have participated either actively or passively during this activity.

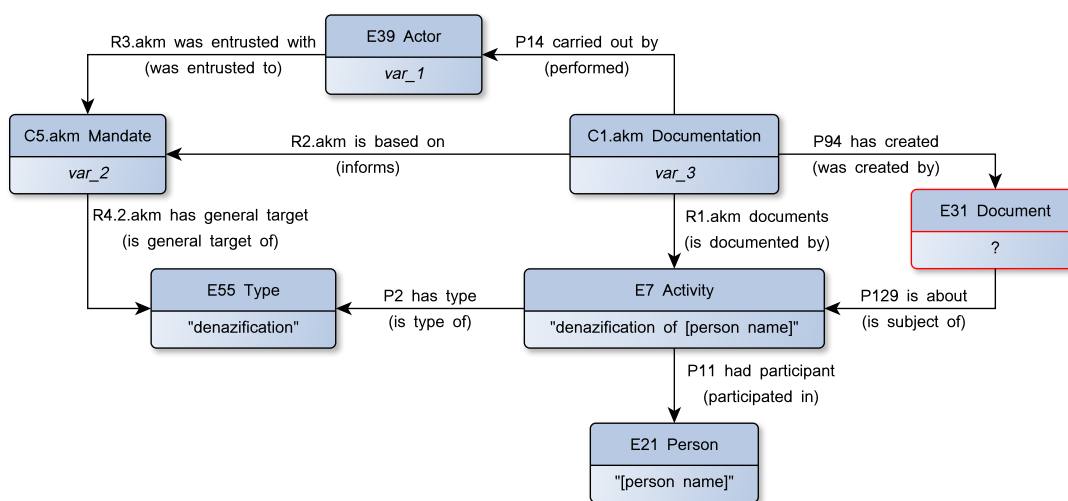


Figure 33 – Q054-04-01<sub>barch</sub>: Documents about the denazification of a person.

Participation, incidentally, does not necessarily entail the person being physically present at any time during this activity. Furthermore, the particular activity “denazification of [person name]” may have existed of a series of subsequent activities which, however, are subsumed

in one instance of the class *E7 Activity*. The *Events* pattern (V:2.2.2) discusses such issues of representing events and activities in much more detail.

Figure 33 shows the principle query pattern, again subsuming various particular query options. A potential query would target documents about an activity with the type “denazification” and in which a particular person [“person name”] participated. These documents have been further created during a documentation activity which recorded the activity to which the documents refer and which was based on a mandate generally targeting activities of the type “denazification”. In this particular case, the place of residence of the person is important. This gives an indication as to where relevant documents might be found since the local authorities were commonly responsible for carrying out acts of denazification.

The *E39 Actor* carrying out the documentation activity might also be considered the same actor to carry out the activity of denazification. In this case, a self-documentation would be pertinent, which is introduced next as a sub-pattern of the *Documentation* pattern.

**Statistics** In the whole sample, the *Documentation* pattern was assigned to 180 questions representing 38% of all inquiries in the sample. Of these 180 questions 131 occur in the BArch sample, representing 38% of all questions in that sample, while the remaining 49 questions appear in the NAN sample representing 37% of all questions in that sample. Within the *Documentation* pattern 97 questions (54%) occur with “actor” as its primary entity of interest, of which 66 belong to the BArch sample representing 50% of all question assigned to the *Documentation* pattern in that sample, while the remaining 31 questions in the NAN sample represent 63% of *Documentation* pattern questions in that sample. Equally relevant are particular groups with 69 counts (38%), of which most appear in the BArch sample with 57 questions (44%) and 12 questions (24%) in the NAN sample. Particular persons are much less frequent with only 10 counts (5.6%) in the *Documentation* pattern of which four (3.1%) belong to the BArch sample and six (12%) to the NAN sample. Types are insignificant: “type of person” does not occur while “type of group” is assigned only four times in the BArch sample (3.1%). The figures indicate that the fact that activities were documented is of importance for many inquiries, often without the specification of a particular actor. However, particular groups and their documentation activities are also relevant in the context of many inquiries.

<i>Provenance Context</i>	<i>All</i> (n=467)		<i>BArch</i> (n=345)		<i>NAN</i> (n=131)	
<b>Documentation</b>	<b>180</b>	<b>38%</b>	131	38%	49	37%
- Documentation [ actor ]	97	54%	66	50%	31	63%
- Documentation [ particular person ]	10	5.6%	4	3.1%	6	12%
- Documentation [ type of person ]	0	-	0	-	0	-
- Documentation [ particular group ]	69	38%	57	44%	12	24%
- Documentation [ type of group ]	4	2.2%	4	3.1%	0	-

Table 6 – Occurrences of the general pattern *Documentation*. Note that percentages for the primary entities of interest are per current general pattern.

**Summary** The *Documentation* pattern covers acts of direct or indirect documentation of unintentional events or of the activities of actors. These documentation acts result in documents that are about the observed events or activities. The differentiation between the act of documenting and the documented act as well as between the actor conducting the observational act and the actor who is being observed is fundamental to the semantics of the pattern. The class  $C1_{akm}$  *Documentation* and the property  $R1_{akm}$  *documents (is documented by)* have been introduced in order to distinctively represent who conducted an act of documentation and who conducted the documented act. Accordingly, the place and time of both acts and the documents which may result from both acts can be specifically modelled.

Furthermore, by representing a distinct act of documentation their purposeful conduct based on a mandate can be represented by the new class  $C5_{akm}$  *Mandate* and the new property  $R2_{akm}$  *is based on (informs)*. The mandate may further specify either the principal or specific target of a mandate via the new properties  $R4.2_{akm}$  *has general target (is general target of)* and  $R4.1_{akm}$  *has specific target (is specific target of)*. Mandates play a pivotal role in archival documentation and have been identified here, based on empirical data, as central to inquiries by many users. Finally, the actor tasked with carrying out the documentation can be specified by the new  $R3_{akm}$  *was entrusted with (was entrusted to)*.

The *Documentation* pattern resembles the creation context for the archival concept of a record, of which an important characteristic has been described as having been “created by persons or devices participating in or observing the activity represented in the record” (Anderson, 2013, 365). This observational aspect is embodied in the documentation activity.

The pattern addresses, especially in combination with the mandate, questions about who might have written – that is, observed and documented – the activities of actors and thus covers a broad range of an essential type of historical question.

Documentation of one’s own activities is a special and distinct case of documentation activities and will be discussed next in the context of the sub-pattern *Self-Documentation*.

### 2.1.3 Self-Documentation

While documentation activities cover the observation and documentation by one actor of another’s activities, many such documentation activities are in fact the observation and documentation of one’s own. One of the most prominent examples would be a group documenting its own meeting by means of minutes.

**General Pattern** The *Self-Documentation* pattern covers the *context of creation* of documents as the result of the direct or indirect observation and documentation of an actor’s own activities. This pattern is a specialization of the *Documentation (V:2.1.2)* pattern. The *Self-Documentation* pattern exhibits one distinct principal sense:

1. The documents (E31) created by a direct or indirect documentation or observation of one's own activities: For example, a parliamentary group documents its own meetings, or travelers on a journey write a report about their trip.

Figure 34 shows the general *Self-Documentation* pattern. Self-documentation is represented with the new class  $C2_{akm}$  *Self-Documentation* as a sub-class of  $C1_{akm}$  *Documentation*. The main difference to the parent pattern is that instances of the class  $E7$  *Activity* are documented that are carried out by the *same* instance of the class  $E39$  *Actor* who is also carrying out the self-documentation.

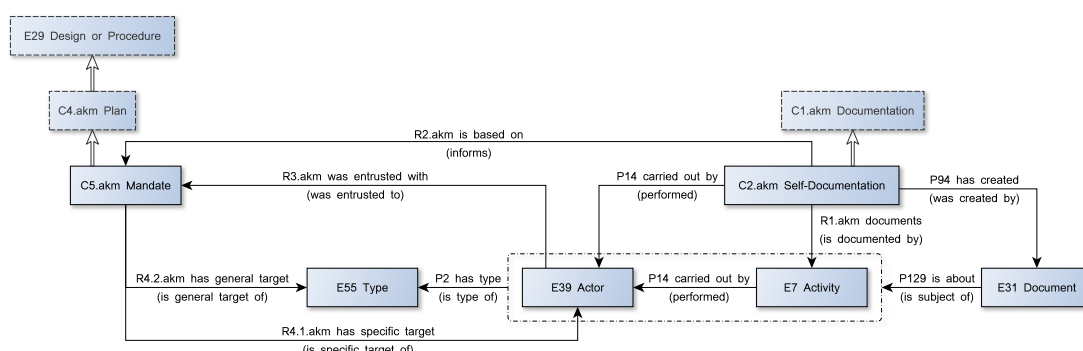


Figure 34 – General *Self-Documentation* pattern.

Figure 2.1.3 shows that an instance of the class  $E39$  *Actor* carries out both the self-documentation activity ( $C2_{akm}$ ) and the activity ( $E7$ ) itself, which is documented by the self-documentation activity.

As in the case of the *Documentation* pattern, the self-documentation is based on a mandate which determines the self-documentation in terms of its principle targets ( $R4.1_{akm}$  and  $R4.2_{akm}$ ). These targets are either types of activities or particular activities, both of which are, again, carried out by the same instance of  $E39$  *Actor* who has been entrusted ( $R3_{akm}$ ) with the mandate ( $C5_{akm}$ ) to document these activities.

The result of such mandate-based self-documentation are often *official accounts* (*Rechenschaftsberichte*) such as proceedings, government statements, or minutes, but also other “unofficial” documents on the internal business and proceedings of agencies.

In other words, the viewpoint potentially found in such documents pertains to an actor and what this actor would like others to know or believe about his or her own activities. This viewpoint can be an internal one, in this case potentially relatively “unfiltered” and typically “non-public”, or a public one, in which case the documented perspective is likely to be filtered.

For example, the members of a parliamentary committee document their meetings by taking minutes. Such meetings are carried out by all participants of that meeting and every participant is formally a contributor to its minutes.

The class *E65 Creation*, of which *C2<sub>akm</sub> Self-Documentation* is a sub-class, refers to those actors who created the content of a document and not simply the person who actually wrote down the text. This difference is not explicitly addressed in the general pattern; however, the CRM property *P14.1 in the role of* would permit the specification of the role played by an actor during a creation activity. Large quantities of documents in state archives are probably the results of such self-documentation activities; that is, activities where an administrative body or state institution created documentation about its everyday business.

### Example - “Minutes of a particular group”

Context                    *“I would like to visit the Bundesarchiv SAPMO in Berlin in order to look up various documents of the KPD and SPD for the time between 1914 and 1933.”*

Q002-01-04<sub>barch</sub>    *“[I would like to see] the minutes of the Reichstag (1919-1930).”*

The general question type is *resource discovery <specific type>*. The wanted entity is *<activity>*. The relevant *given* entities are *particular group* (“Reichstag”), a *time-span* (“1919-1930”), and *type of document* (“minutes”). The *supplementary* information includes *particular group* (“KPD”, “SED”).

The question is a *material fact* question since it inquires about an observable relationship between a document, the minutes, and documented activities. The *provenance context* is *self-doc [particular group]* since we are looking for documents the Reichstag created about its own activities. The *aboutness context* is *events [type of activity]*.

The question does not specify any particular activity or type of activity documented by the requested minutes. However, since the user is asking for minutes of the Reichstag, “meeting” as a general type of activity can be inferred. Similarly, since a general type of requested document is given in the question, the query could specify a controlled term such as “minutes” as the type of the requested documents.

Figure 35 shows a possible query instantiation: The query asks for any documents (E31) of the type “minutes” which are about any activity (E7) of the type “meeting”. These activities occurred during the time-span (E52) 1919 - 1930. The documents are the result of any self-documentation (C2<sub>akm</sub>) activity recording any activities of the type “meeting”, both of which have been carried by the same actor (E39). This actor was a member of the Reichstag (E74) and had been entrusted with a mandate (C5<sub>akm</sub>), here the particular example of the “parliamentary rules of procedure of the Reichstag”, to take minutes of meetings on which the self-documentation activity was based.

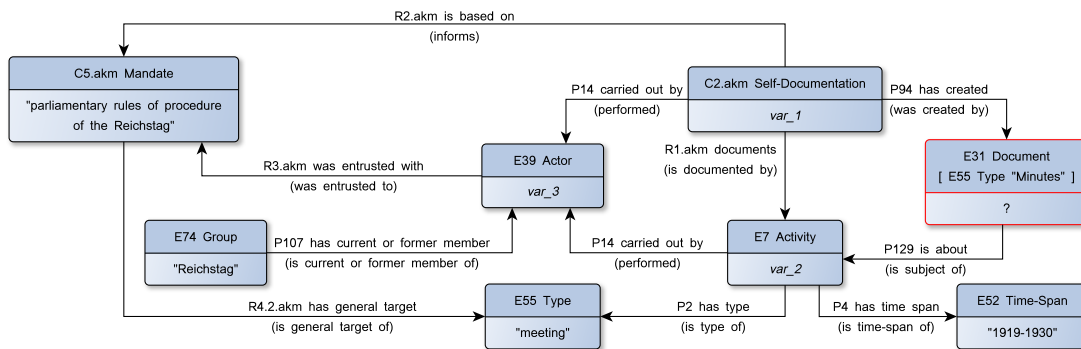


Figure 35 – Q002-01-04<sub>barch</sub>: Minutes of the Reichstag.

Such a query would retrieve any documents of the type “minutes” documenting any “meeting” activity of an actor who was a member of the Reichstag and who documented these meetings according to the particular mandate “parliamentary rules of procedure of the Reichstag”.

In this query pattern, the generic class *E39 Actor* is used because the user’s inquiry does not specify whether only particular groups or type of groups conducted such self-documentation activities or whether particular persons or type of persons also did so. By using the generic class *E39 Actor* the query remains inclusive in this regard.

Similarly, by using the fictitious term “meeting” as an example for the type (E55) of activity (E7), any such activity would be included in a query. Controlled terms from a controlled vocabulary would allow an easy adaptation of the query to target more specific types of activities of the Reichstag such as parliamentary or committee meetings. The latter is the case in the next example.

### Example - “Minutes of committee meetings”

Context “*[person name]*, as a delegate in the Reichstag from 1920 until 1930, was active in various committees of the Reichstag.”

Q002-05-02<sub>barch</sub> “Are the minutes of these committee meetings (...) in the Bundesarchiv?”

The general question type is *resource discovery* <specific type>. The wanted entity is <activity>. The relevant given entities are *particular person* (“*[person name]*”), *particular group* (“Reichstag”), *type of group* (“committees of the Reichstag”), *time-span* (“1920-1930”), the *type of activity* (“committee meeting”), and *type of document* (“minutes”).

The question is a *material fact* question since it inquires about an observable relationship between documents and a type of activity. The *provenance context* is *self-doc* [*type of group*] since we are looking for documents the Reichstag created about its own activities. The *aboutness context* is *events* [*type of activity*].

While this question is very similar to the previous one, here the user is more specific regarding the requested minutes. In this example, a specific type of meeting is given, “committee meetings”, and a particular person is given who has been active in various committees of the Reichstag. The statement “was active” can be interpreted as the membership of that particular person in various committee groups of the Reichstag, such as particular committees on legal or financial affairs. The user is interested in the minutes of activities which were of the type “committee meeting” and have been carried out by those groups of which the particular person was a member.

Figure 36 shows a principal query pattern which mostly resembles that of Figure 35.

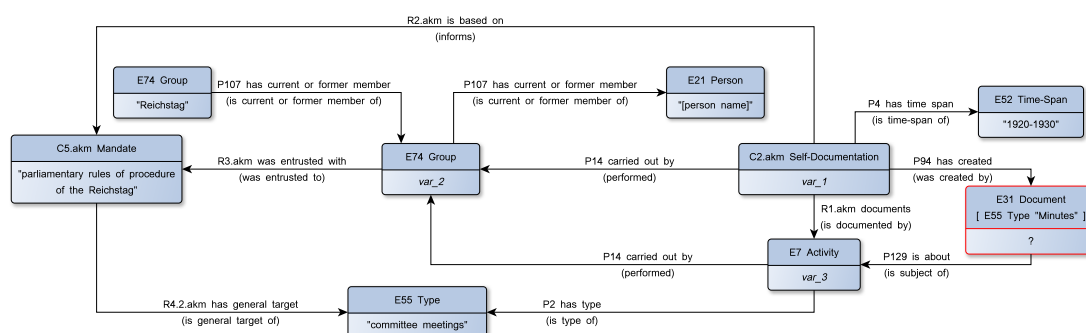


Figure 36 – Q002-05-02<sub>barch</sub>: Minutes of committee meetings.

The query pattern could be modified to retrieve additional potentially relevant documents. For example, the question could be interpreted as focusing on committee meetings in which the particular person participated regardless of whether he or she was a member of a particular group. In this case, the entity *E74 Group* could simply be replaced with the entity *E21 Person* in Figure 36.

### Example - “Report of a particular activity”

Context *“In 1980, a delegation of the FDGB led by Harry Tisch laid down a wreath of flowers in Oradour. The visit was part of a trip of the FDGB to France (...). At this time, Tisch was also a member of the Politbüro of the ZK of the SED.”*

Q005-08-02<sub>barch</sub> *“Where can (...) the report on this trip (...) be found?”*

The general question type is *resource discovery <specific item>*. The *wanted* entity is *<activity>*. The relevant *given* entities are *particular group* (“FDGB”, “Politbüro of the ZK”, “SED”), *particular person* (“Harry Tisch”), *particular place* (“Oradour”, “France”), *time-span* (“1980”), *particular activity* (“trip to France”, “laying down of a wreath of flowers”), and *particular document* (“report”).

This is a *material fact* question since it inquires about an observable relationship between a document and its topic. The *provenance context* is *self-doc [actor]* since an actor who must



have participated in the trip created the report; at least the context of the inquiry suggests that an official or internal report is requested and not other types of account such as newspaper articles, in which case a documentation activity would be prudent. The *aboutness context* is *events [particular activity]*.

The provenance context of this question is specified as self-documentation because a report about an activity in the sense of a direct account can only be written by someone who has participated in the activity. The query pattern therefore assumes an unknown person (E21) to have carried out both the self-documentation and the documented activity.<sup>74</sup>

The further context of the question leads us to conclude that the user is looking for an official report on the trip to France rather than any kind of indirect or unofficial account. The exemplary query pattern therefore includes as an additional parameter an unknown mandate which has been entrusted to the person who carried out the self-documentation, in turn based on this mandate. The query is thus limited to official reports of the trip insofar as these reports have resulted from self-documentation activities based on an official mandate.

For the requested documents, the type “report” may be additionally assumed. The usefulness of this information depends on the existence of an adequate controlled vocabulary listing appropriate terms for types of documents. The type of document may also be derived from the particular or the type of mandate or self-documentation. Even though the user asks for the specific item “*the*” report on the trip to France, there may of course be several reports available. The query, therefore, inquires after any document reporting on the trip to France as the result of the described provenance context.

Figure 37 shows a possible query pattern. The query asks for any document (E31) of the type “report” (E55) which is about the particular activity (E7), “the trip to France of the FDGB in 1980”, and which has resulted from any kind of self-documentation activity (C2<sub>akm</sub>). Both the self-documentation and the trip to France must have been carried out by the same unknown person (E21) who, at the same time, had been entrusted with a mandate (C5<sub>akm</sub>) to write a report (C2<sub>akm</sub>) on the trip (E7).

---

<sup>74</sup> Note that “report” could be interpreted differently as also including indirect accounts of the trip such as a newspaper article. In this case, the documentation activity could be used in order to widen the scope of the query.

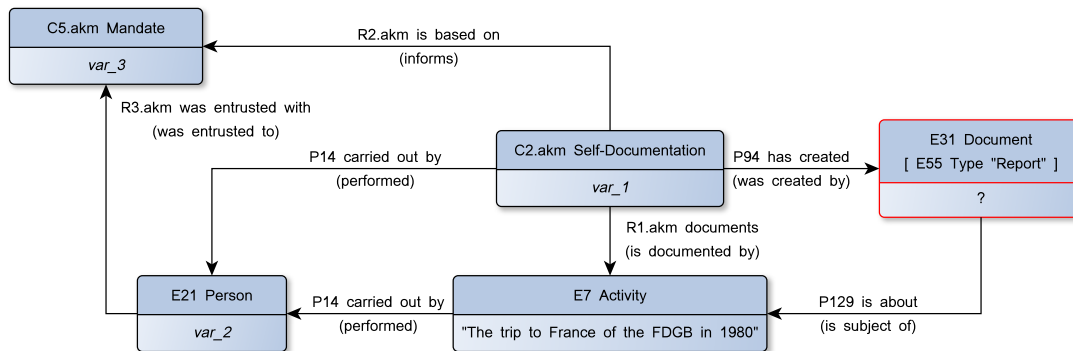


Figure 37 – Q005-08-02<sub>barch</sub>: Report on a trip to France.

The trip to France of the FDGB in 1980 most likely occurred before the actual writing of the report. The activity of self-documentation in this example conceptually comprises the extent of the documented activity, the trip to France, and the time-span during which the report was written.

The particular activity “the trip to France of the FDGB in 1980” could be more explicitly described by additional given contextual information, such as the place “France”, the date “1980”, or the group “FDGB” as the actor carrying out the activity. The role of contextual information for specifying events and activities is described in detail in the *Events* pattern (V:2.2.2).

Since the user provided additional contextual information in the inquiry, the rather general query pattern shown in Figure 37 could be further specified with particular given information. For example, instead of querying for any person, the query could specifically ask for the particular person, “Harry Tisch”, as the actor who participated in the trip to France and wrote a report of it.

On the other hand, a query could try the particular group “FDGB” as the actor who carried out both activities. Finally, by using the class *E39 Actor*, the query would remain inclusive regarding whether a particular group or person acted in the context of the two activities. In this regard, the question is a good example for possible indirections.

**Statistics** In the whole sample, the *Self-Documentation* pattern has been assigned to 49 questions representing 10% of all questions, of which 39 belong to the sample from the BArch representing 11% of all question in that sample. The remaining 10 questions belong to the sample from the NAN and represent 7.6% of all questions in that sample.

The most important primary entity of interest are particular groups with 26 counts in the whole sample, representing 53% of all inquiries. Of these 26 most belong to the BArch sample with 22 questions which represent 56% of all inquiries assigned to the *Self-Documentation* pattern in that sample. Only 4 self-documentation questions (40%) with a particular assigned group appear in the NAN sample. In the context of self-documentation, only one particular person

(2%) appears in the whole sample from the NAN. Types of group occur four times (8.2%) in the whole sample of which three (7.7%) are part of the BArch sample and one (10%) of the NAN sample. “Type of persons” does not occur.

<i>Provenance Context</i>	<i>All (n=467)</i>		<i>BArch (n=345)</i>		<i>NAN (n=131)</i>	
<b>Self-Documentation</b>	<b>49</b>	<b>10%</b>	39	11%	10	7.6%
- Self-Documentation [ actor ]	18	37%	14	36%	4	40%
- Self-Documentation [ particular person ]	1	2%	0	-	1	10%
- Self-Documentation [ type of person ]	0	-	0	-	0	-
- Self-Documentation [ particular group ]	26	53%	22	56%	4	40%
- Self-Documentation [ type of group ]	4	8.2%	3	7.7%	1	10%

Table 7 – Occurrences of the general pattern *Self-Documentation*. Note that percentages for the primary entities of interest are per current general pattern.

In many cases, the more general *Documentation* pattern (V:2.1.2) has been assigned even though the more specific *Self-Documentation* pattern would also have been applicable. The numbers indicate that the self-documentation of particular groups is relevant; that is, documents about one’s own activities such as reports or minutes.

**Summary** The *Self-Documentation* pattern extends the *Documentation* pattern and specifically describes the documentation of one’s own activities. This idea of observing one’s own activities, again in a direct or indirect manner and possibly based on a mandate, is represented by the new class *C2<sub>akm</sub> Self-Documentation*.

The pattern emphasizes official or unofficial, public or internal accounts of an actor, especially of state and administrative agencies, regarding the activities of this actor and what this actor would like other actors to know or believe regarding these activities. The decisive criterion is that an actor presents his or her own viewpoint regarding his or her own actions. The designation – internal or external and thus public – of such self-documents can be modelled using the *Correspondence* pattern (91:162).

In particular, “internal” documents, for example, minutes or accountability reports, are of high significance in state archives. Their value for historical research is that they result from routine and the daily business of administrative bodies and state agencies, thus documenting their plans, decisions and discussions, from internal and potentially “unofficial” and relatively “unfiltered” perspectives. Questions about how and why particular groups reached decisions or inquiries focusing on the reasons, plans (V:2.2.3) and aims of particular groups will utilize this pattern.

#### 2.1.4 Correspondence

The information objects in the archive can be understood as recorded acts of communication. They are created typically *for* someone, they are addressed to someone, and they are taken notice of or received by someone.

**General Pattern** The *Correspondence* pattern covers the *context of creation* of things that have been designated for, or intended for receipt by, another actor. The pattern serves questions about recorded acts of communication. Three main senses can be distinguished:

1. The creation activity (E65) which creates identifiable but *immaterial* things (E73) and is *designated* for another actor. The designation of a creation activity may be known by history scholars or given by a mandate (C2<sub>akm</sub>); for example, the writing of a personal letter to a friend, the petitioning to a government, the sending of a diplomatic communiqué, or the delivery of a speech.
2. The identifiable but *immaterial* things (E73) which have been *addressed* to an actor. This phenomenological knowledge is derived from written evidence such as the address line on a document, recipient notes on internal documents, or the recordings or manuscripts of speeches.
3. The *persistent* and *tangible* things (E24) of which an actor *took notice* insofar as the actor kept the thing at some point, such as a personal letter received from a friend, internal working documents circulated between departments, or a message intercepted by a secret agency.

The activity of sending or distributing a document as an activity independent from the creation of that document could be considered another relevant sense. For example, leaflets or pamphlets are distributed by actors who did not create these documents. The pattern does not distinguish between someone who commissioned the creation of the document, those who actually created the document or contributed to the creation, and those who dispatched or otherwise distributed the documents. The necessity for such a level of complexity was not supported by the research data and appears relevant only for very few specialized questions. Specific activities for commissioning, sending or distributing are therefore not introduced in the *Correspondence* pattern. Figure 38 shows the general *Correspondence* pattern.

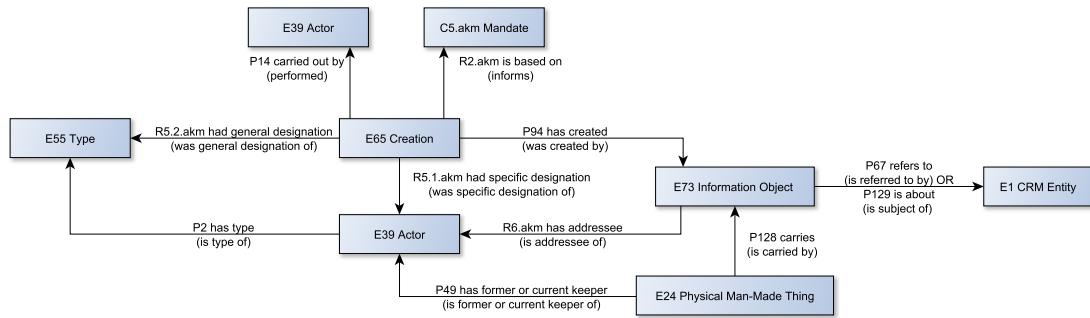


Figure 38 – General Correspondence pattern.

The first relevant historical situation describes creation activities (E65) designated for an actor. In other words, the information objects are created with the intention to address or reach someone else; for example, writing (E65) a letter with the intention to send it to another person, or giving (E65) a radio speech addressing an audience. Whether the letter has been sent or arrived at its intended destination or whether the radio speech has been heard or understood does not matter in this context.

The new property *R5.1<sub>akm</sub> had specific designation (was specific designation of)* designates the intended particular recipient or audience of the creation activity; for example, the mayor of Berlin or the Reichstag. The new property *R5.2<sub>akm</sub> had general designation (was general designation of)* expresses the intended principal type of actor as the designation of a creation activity; for example, privately owned companies or lawyers. The material fact that a creation activity had been designated for an actor is derived from historical background knowledge.

This creation activity may be based on a mandate (C5<sub>akm</sub>) as discussed in the context of the *Documentation* pattern (V:2.1.2), which may also permit the derivation of the designated actor; for example, the right of members of the Reichstag to address the plenum based on parliamentary law (C2<sub>akm</sub>).

Furthermore, delivering a speech and thus creating an information object, the contents of the speech, what is said, and which may be recorded (C1<sub>akm</sub>) by stenographers or audio devices, also points towards certain actors as those designated to receive said speech.

As already discussed in the *Provenance* pattern (V:2.1.1), the creation (E65) of information objects (E73) takes place at (P7) geographical places (E53), happens (P4) during a time-span (E52), and may fall into (P10) a period (E4).

Closely related to the intention and designation is the phenomenological fact of an information object being addressed to someone. The new property *R6<sub>akm</sub> has addressee (is addressee of)* relates an instance of the class *E73 Information Object* with an instance of the class *E39 Actor*. This material fact allows for the representation of the phenomenological fact of the addressee of an information object as given somewhere on the document. For example, the address line on a letter or the salutation in a speech may provide the intended recipient. The instance of *E73*

*Information Object* may be about any instance of *E1 CRM Entity* as discussed in more detail in the context of the *Aboutness* pattern (V:2.2.1).

Both the historical knowledge about the intended designation of information objects and the phenomenological knowledge about the addressee of information objects are closely related. They do not necessarily entail, however, the message or physical document, for example a front-line report, reaching its destination. Similarly, a radio speech may have been addressed to the German people, but it remains questionable whether all members of the German population did in fact hear, to say nothing of comprehend, the radio speech.

The second historical situation of receiving or otherwise obtaining a document or, even more generally, taking notice of a document, is independent from intended designations or phenomenological addressees. As already discussed in the context of the *Provenance* pattern (V:24), possession or having taken notice of, is represented by the property *P49 has former or current keeper (is former or current keeper of)*. The property allows for the material fact that an actor has been the former or current keeper of a physical document, often a copy of the original. How this actor came into possession of said document, legally or illegally, or if this actor was the addressee or intended recipient, remains unresolved.

Possession of a document does not entail knowledge of the actual contents. The thing which has been kept may never have been read. However, the material fact of having been in possession of the thing provides strong evidence, if not proof, that an actor must have been informed.

Various notations, editing marks, receipt stamps, and other annotations on documents provide plenty evidence on addressees and on who must have taken notice of a particular document. In principle, by reusing the class *E7 Activity* (V:2.2.2), the *Correspondence* pattern would allow for the creation of a detailed history of a document, when and by whom it has been created, sent, edited and received, internally and externally. However, the intended level of detail in this study, based on the immediate evidence of the questions, focuses on a more coarse and aggregated description of such information on a document. In order to identify sets of relevant documents it suffices to have a summary of all the intended addressees and those who had possession of the documents at some point. The analyzed inquiries in the sample do not provide sufficient evidence that more detailed descriptions are necessary in order to discover potentially relevant documents.

### **Example - “Personal Correspondence”**

Context                    *“Predominantly I am interested in the personal estates of the following personages of the literary and cultural scene in the GDR: [person name].”*

Q003-01-02<sub>barch</sub> *“In these holdings I am especially interested in (...) correspondence (...).”*

The general question *type* is *resource discovery <specific type>* since the question asks for a specific type: in this case, “correspondence”. The *wanted* entity is *<person.by>*. The relevant *given* entities are *particular person* (“[person name]”), *particular document* (“personal estate of [person name]”),

“correspondences”). The *supplementary* information includes *particular group* (“the literary and cultural scene in the GDR”), and *particular place* (“GDR”). The question is a *material fact* question. The *provenance context* is *correspondence [particular person]*. The *aboutness context* is *aboutness [any]*.

Figure 39 shows the first principal instantiation of the query pattern. The query pattern asks for all the physical things, reproductions or originals, that carry content created by the particular person in question (“[person name]”) and, at the same time, have been either designated for another actor, in this case unknown, or that carry addressee information pointing to another actor. Since the user is specifically asking for correspondence in the personal estate of “[person name]”, the query could be further limited to physical things from this particular personal estate.

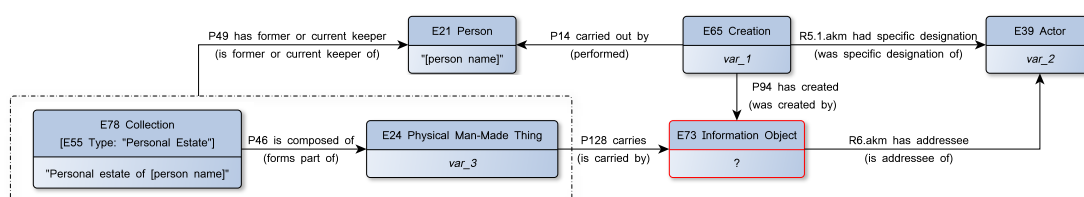


Figure 39 – Q003-01-02a<sub>barch</sub>: Things created by an actor and *designated* for another actor.

Since the user is interested in the correspondence, a query should also target any information objects the person in question received. Figure 40 shows the second principal instantiation. Relevant here are information objects that have been designated for the person in question during an creation activity, or information objects that are addressed to the person in question. Since the user is interested only in information objects from the personal estate, the second principle query also includes the physical carrier (E24) of which the person in question must have been the keeper, or which must be part of the personal estate (E78).

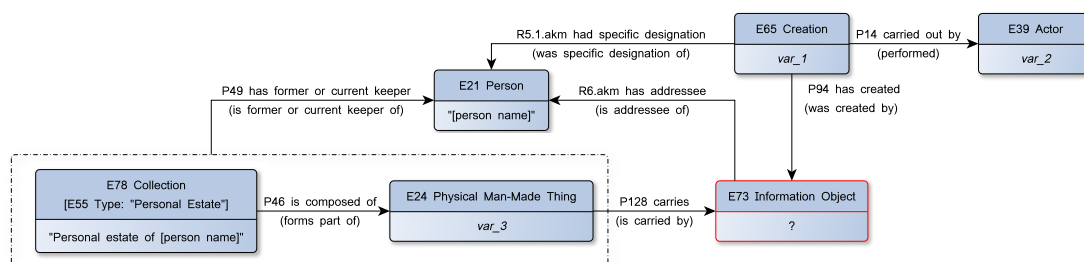


Figure 40 – Q003-01-02b<sub>barch</sub>: Things created by an actor and *designated* for another actor.

Again, receipt, or taking notice, of an information object is indicated by the material fact that the person has been the keeper of a physical carrier carrying the information object designated for and sent to this person. Of course, a query could inquire after any information objects

addressed to or designated for the person in question that are stored in other holdings and personal estates; for example, copies of letters made before sending the letter. In this case, the property P49 in Figure 40 would point to the class E39 on the right hand side and originate from the class E24. Then, however, the question as to whether the addressee actually acknowledged the letter, having received it, would remain open and be subject to further critical investigation by the historian.

### Example - “Contact with a Group”

Context                    “[I would like to see] documents of the ‘Büro Krenz’ about the opening of the border in November 1989, ...”

Q009-01-11<sub>barch</sub> “...especially relating to contacts with Czech and Soviet authorities (...).”

This example is a twofold question which first asks for any documents of the “Büro Krenz” about the opening of the border in November 1989 and then, secondly, specifically for those documents which originate from contacts with Czech and Soviet authorities. The latter sense will be taken as the example here.

The general question type is *resource discovery* <material-finding>. The *wanted* entity is <activity>. The relevant *given* entities are *particular group* (“Büro Krenz”), *type of group* (“Czech authorities”, “Soviet authorities”), *particular activity* (“opening of the border in November 1989”), *type of activity* (“contacts with Czech and Soviet authorities”), and *time-span* (“November 1989”). The question is a *material fact* question. The *provenance context* is *correspondence* [particular group] since the particular group “Büro Krenz” and its correspondence is the focus of the question. The *aboutness context* is *events* [particular activity].

While the first part of the question on any documents of or by the “Büro Krenz” about the opening of the border in November 1989 can be answered with the *Documentation* (V:2.1.2) and *Self-Documentation* (V:2.1.3) pattern, the second part of the question is more specific and requires the *Correspondence* pattern.

The interpretation of the second part of the question is that the user seeks any original document or copies of documents about the border opening in November 1989, which (1) were created by the Büro Krenz and designated for Czech and Soviet authorities, regardless of whether these documents were actually sent or even arrived at their destination(s), or whether they remained drafts, or which (2) were received by the Büro Krenz from Czech and Soviet authorities, which would mean that the “Büro Krenz” is the former or current keeper of such documents.

It is also possible that, once again, minutes about the creation event might be relevant; for example, meetings discussing the communication with Czech and Soviet Authorities. This case would be covered by the *Documentation* (V:2.1.2) pattern. Here, the focus rests on the first interpretation which is shown in Figure 41 and the second interpretation, which is shown in Figure 42.



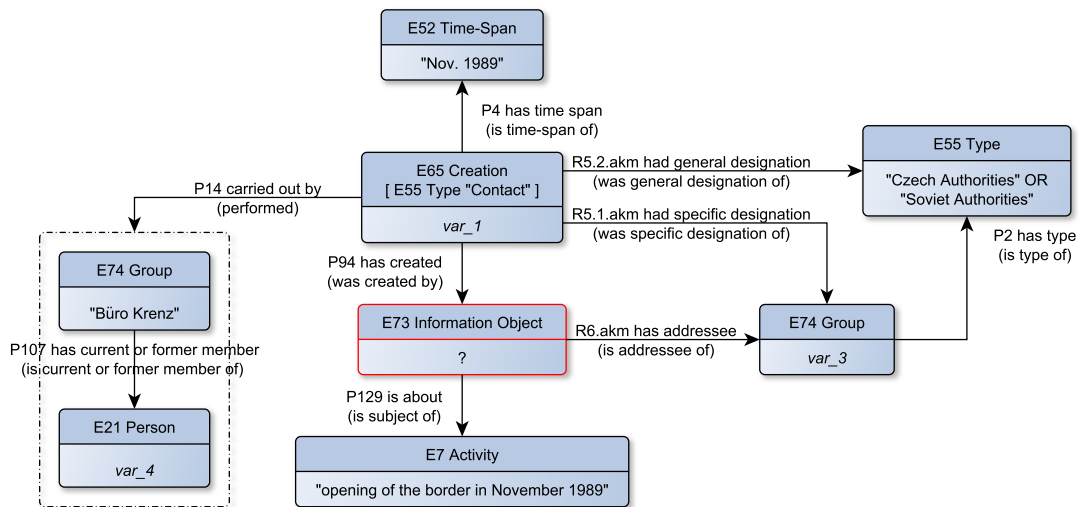


Figure 41 – Q009-01-11a<sub>barch</sub>: Documents as the products of *contacts* with authorities.

The group “Büro Krenz” or any member of this group carried out an creation activity during November 1989, such as formulating and writing a communiqué. This creation activity could very well have been a meeting of members of the “Büro Krenz” and documented by minutes in a self-documentation activity.

Here, however, the decisive criterion is that this creation activity occurred, at least partly, with a type of group as the designated recipient; in this case, the Czech or Soviet Authorities, and created an information object for this group. This information object, the contents of a communiqué, for example, may indicate an addressee, and is about the particular activity of “opening of the border in November 1989”.

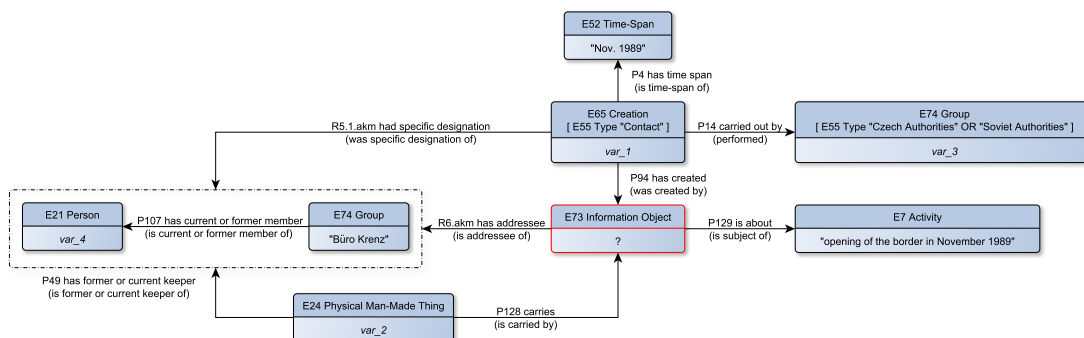


Figure 42 – Q009-01-11b<sub>barch</sub>: Documents as the products of *contacts* with authorities.

In fact, all information objects are potentially relevant whose contents have been created by either of the two types of groups and address the particular activity “opening of the border

in November 1989”, and which have further been either designated for or addressed to the particular group “Büro Krenz”, or one of its members.

### Example - “Petitions of the Population”

Q039-01-03<sub>barch</sub> “A special emphasis [of my work] is on the role of the social policy of the SED leadership. (...) One focal point is the aim of the SED to solve the housing problem. Here I would like to see the petitions of the population.”

The general question type is *resource discovery* <specific type>. The wanted entity is <group.by>. The relevant given entities are *particular group* (“SED”), *type of group* (“population”), *type of activity* (“social policy”, “aim to solve the housing problem”), and *type of document* (“petitions”). The question is a *material fact* question. The *provenance context* is *correspondence* <type of group>. The *aboutness context* is *plans* <type of activity>.

The interpretation of this question is that the user would like to see petitions or similar types of documents created by members of the populace of the GDR designated to the SED leadership. Further, these documents must generally be about or related to the “social policy of the SED”. Figure 43 shows how this interpretation can be represented using the *Correspondence* pattern. The query pattern assumes a creation activity representing the many activities of the type “petitioning” carried out by actors, most likely individual persons but potentially also groups, who were members of the particular group “population of the GDR”.

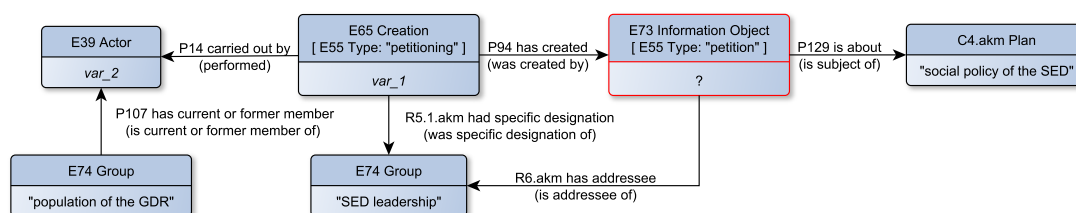


Figure 43 – Q039-01-03<sub>barch</sub>: Petitions as special acts of direct communication.

The creation activities were designated for the particular group “SED Leadership” and have created information objects of the type “petition” which are about the plan “social policy of the SED”. The information objects are addressed to the group “SED Leadership”. The class *E24 Physical Man-Made Thing* is not needed in this query pattern because the question as to whether or not the SED actually received the petitions is immaterial in this case.

**Statistics** In the complete sample, the *Correspondence* pattern is assigned 21 times, which represents 4.4% of all questions. Of these 21 questions, only 2 belong to the NAN sample, representing 1.5% of all questions in that sample while the remaining 19 represent 5.5% of all question in the BArch sample. Two *Correspondence* patterns in the NAN sample pertain to

“actor”. In the BArch sample, seven questions (37%) refer to a particular person, five to a type of group (26%), four to a particular group (21%), and three to an actor (16%).

<i>Provenance Context</i>	<i>All</i> (n=467)		<i>BArch</i> (n=345)		<i>NAN</i> (n=131)	
<b>Correspondence</b>	<b>21</b>	<b>4.4%</b>	<b>19</b>	<b>5.5%</b>	<b>2</b>	<b>1.5%</b>
- Correspondence [ actor ]	5	24%	3	16%	2	100%
- Correspondence [ particular person ]	7	33%	7	37%	0	-
- Correspondence [ type of person ]	0	-	0	-	0	-
- Correspondence [ particular group ]	4	19%	4	21%	0	-
- Correspondence [ type of group ]	5	24%	5	26%	0	-

*Table 8* – Occurrences of the general pattern *Correspondence*. Note that percentages for the primary entities of interest are per current general pattern.

Even though the *Correspondence* pattern occurs only in relatively few cases, it offers fundamental relationships pertaining to the historical contexts of most information objects; that is, information being intended for, addressed to and received by an actor. The general pattern therefore may be considered supplementary in many other cases which have been not counted separately, especially for psychological and collective statistical questions. As in the example questions on pages 164 and 166, this pattern may apply twice or even more often since questions are typically interested in the correspondence between two or even more actors. However, the pattern is counted only once per question.

**Summary** The core semantic of the *Correspondence* pattern covers three essential aspects of communication acts: (1) the intended or unintended *designation* of information for a particular or a type of actor through a written or oral creation activity which may, further, be based on a mandate; (2) the phenomenological knowledge regarding the *addressee* of information, and (3) the notion of *taking notice* of information by having kept and thus received it in some form or another.

Three new properties have been introduced in order to explicitly represent the aforementioned semantics of *designation* and *addressing*. The property *R5.1<sub>akm</sub> had specific designation (was specific designation of)* expresses that an information object has been specifically designated to a particular actor (E39), while the property *R5.2<sub>akm</sub> had general designation (was general designation of)* expresses an principal designation to a type of actor (E55). Both properties are good examples for the utilization of implicit background knowledge of the archivist during documentation about the particular historical context described. Such specialized knowledge is necessary in order to assert such material facts. Necessary evidence for designation may be further derived from mandates (C5<sub>akm</sub>) – a good example of how the general patterns overlap and “reuse” entities – or phenomenological knowledge such as the stated or noted addressee of an information

object represented by the new property  $R6_{akm}$  *has addressee (is addressee of)*. Finally, the notion of taking notice is implicitly evident from the material fact that an actor has been in possession of or kept (P49) the piece of information in question at some point during his or her lifetime.

These three basic statements allow us to address a range of basic questions regarding recorded acts of communication such as who was the intended recipient of information, who might have taken notice of it, or who must have had knowledge of that information? In other words: the material facts which can be created based on the *Correspondence* pattern support the process of answering questions pertaining to the principal addressee or intended recipient of communication acts, or the question as to whether someone must have received the information, thus knowing of particular activities or incidents.

In the context of this analysis, this general pattern is assigned to inquiries if they demand the retrieval of recorded acts of communication in order to satisfy their perceived primary interest.

## 2.2 Aboutness Context

The general patterns from the group *aboutness context* focus on the historical context to which an information object or a requested fact is referring. In this regard, the aboutness context also covers the general aboutness of a question but primarily refers to the content reference of documents and facts.

The general patterns that belong to the aboutness context are listed in Table 9 and generally address the principal questions regarding what documents are about, what they describe or discuss, and how they relate to the historical reality. In terms of factual questions, they generally describe what kinds of activities happened, who carried them out, what actors did, planned, wanted or executed.

<i>Pattern Name</i>	<i>Informal Scope Note</i>	<i>Page</i>
Aboutness	A generic, yet inclusive general pattern focusing on the aboutness of resources, facts, and questions. Together with the <i>Provenance</i> pattern the upmost layer of the AKM.	171
- Events	Describes the context of events and activities. Sub-pattern of the <i>Aboutness</i> pattern.	175
— Plans	Describes the context of plans. Sub-pattern of the <i>Events</i> pattern.	184
- Actors	Describes the context of actors. Sub-pattern of the <i>Aboutness</i> pattern.	202
- Things	Describes the context of physical things. Sub-pattern of the <i>Aboutness</i> pattern.	210
— Documents	Describes the context of documents. Sub-pattern of the <i>Things</i> pattern.	215

Table 9 – General patterns belonging to the provenance context.

## 2.2.1 Aboutness

All information objects stored in an archive make statements about physical or conceptual entities in the world. They all have a subject matter and refer to something.

**General Pattern** The pattern *Aboutness* forms, together with the pattern *Provenance* (V:2.1.1), the upmost layer of the AKM in terms of its semantics. All other general patterns are specializations of one of these two patterns.

The pattern *Aboutness* covers the generic *context of aboutness* of information objects. The principle semantic of the pattern is:

1. The things (E1) about which identifiable but *immaterial* things (E73) make a general statement about or have as their primary subject.

Figure 44 shows the core of the principal *Aboutness* pattern: Any instance of the class *E73 Information Object* may make a statement about any instance of the class *E1 CRM Entity*.

The class *E73 Information Object* has been introduced in the context of the *Provenance* pattern (V:2.1.1). The class *E1 CRM Entity* is the topmost concept in the CRM and comprises all things in the discourse universe of the CRM. In principle, therefore, the pattern *Aboutness* allows us to assert any aboutness relationship between an instance of *E73 Information Object* and any instance of *E1 CRM Entity* since the latter is the super-class of all other classes in the CRM.<sup>75</sup>

The nature of this relationship may be either one of *general reference*, expressed by the property *P67 refers to (is referred to by)*, or one of *primary subject*, expressed by the property *P126 is about (is subject of)* which is a sub-property of *P67 refers to (is referred to by)*. For example, a census list (E31) generally refers to (P67) a group of people (E74), while a report (E31) has as its primary topic (P129) the activity (E7) on which it is reporting. The *Aboutness* pattern thus resembles a simple subject relationship between an information object and its subject matter.

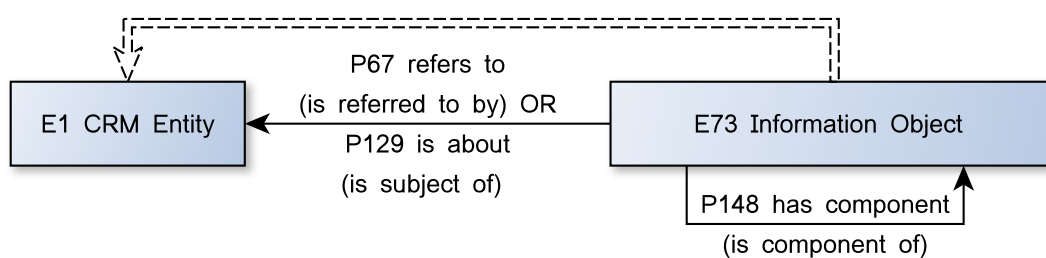


Figure 44 – Core of the *Aboutness* pattern.

Within the scope of this study, the discourse universe is mostly limited to a particular range of relevant entities. The most relevant entities are summarized in Figure 45.

<sup>75</sup> With the exception of primitive values (E59), to be exact.

In resource-discovery questions, users mostly ask for instances of the class *E73 Information Object* which are primarily about an instance of either the class *E7 Activity* or the class *E39 Actor* and their sub-classes. These information entities are given as the qualification of the pattern assignment. The *type of question*, incidentally; that is, resource discovery, indicates that the question is about an information object, that the pattern assignment provides the general aboutness of the question in terms of the historical context to which it refers, and that the information entity given with the pattern assignment indicates the aboutness of the information object requested by the user in a question.

The other patterns, which are specializations of the *Aboutness* pattern, further describe the specific historical context for these classes. This historical context is derived from the interpretation of the questions. The patterns provide the additional relevant entities, mostly the *given* information from the questions, and adequate context from the historical and archival domain.

The *Events* pattern (V:2.2.2) focuses on questions about the context of activities in general, and the *Plans* pattern (2.2.3) on questions about the planning, and wanting and execution of plans, while the *Actors* pattern (V:2.2.4) focuses on questions about the activities of persons and groups. The *Things* pattern (V:2.2.5) describes the history of objects. Questions about information objects (E73) are covered by the *Documents* pattern (V:2.2.6).

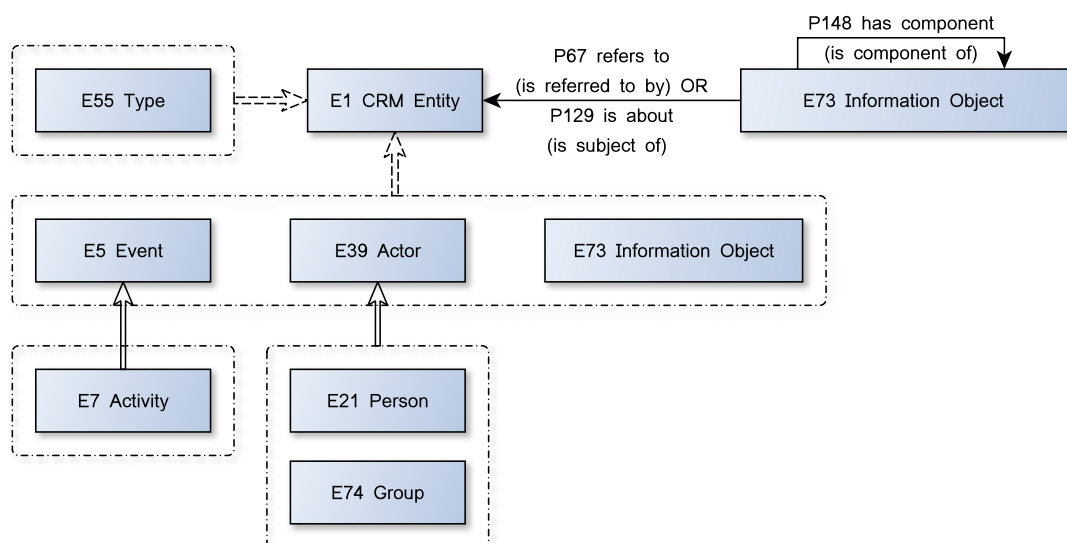


Figure 45 – General *Aboutness* pattern.

Questions referring to information objects (E73) about unintentional events (E5), places (E53), time-spans (E52), periods (E4), or physical things (E24) were not relevant. This means that users did not ask for information objects (E73) focusing on unintentional events such as

accidents (E5) or plagues (E5), the history of particular places such as cities (E53) or countries (E53), physical objects such as monuments (E24) or devotionals (E24), or time-spans (E52) or periods (E4) such as a document about the 1980s.<sup>76</sup>

Proper general subjects are covered by *E55 Type*. The analysis of the questions, however, did not provide any evidence that users inquire after things referring specifically to a general subject.<sup>77</sup> This is not surprising since the inquiries focus on primary documents and facts in order to either describe or gather evidence for historical phenomena. Only one question specifically asked for a secondary source about the “Development of the law of associations and the policy regarding associations 1933-1945”. In principle, however, any primary document may be about or refer to a general subject. The pattern *Aboutness* therefore includes the class *E55 Type*, which may be used to state a general aboutness of an instance of the class *E73 Information Object* in terms of general subject.

Factual questions may inquire about *any* material fact described in the patterns; that is, *Aboutness* and *Provenance* patterns such as the name of a person, the date of joining a political party, or the mandates of a particular group.

Instances of *E31 Document* may have as their primary subject any instance of *E73 Information Object*. This specific case is covered in more detail in the *Documents* pattern (V:2.2.6). For example, a finding aid (E31) or holding guide (E31) describes (P126) the context of archival documents (E73) in a holding. Questions *about* the type or form of a written or visual document or any other qualities of information objects, including features of the archive and its holdings, are also covered by the pattern *Documents* (V:2.2.6).

Since it is assumed that every information object is about something, the *Aboutness* pattern is assigned to a question if the aboutness of the information object to which the question refers is not known or not given, and cannot be reasonably deduced. Since this is seldom the case and only very few questions belong to this category, only one additional example is given here. Other examples for the pattern *Aboutness* include those given in the pattern *Provenance* (V:2.1.1).

### Example - “Reports about anything”

Context                    “I would like to visit the Bundesarchiv SAPMO in Berlin in order to look up various documents of the KPD and the SPD for the period between 1914 and 1933: (...)”

Q002-01-09<sub>barch</sub> “Reports of the ZRK; (...)”

The general question type is *resource discovery <specific type>*. The *wanted* entity is *<group.by>*, more specifically documents of the type “report” created by the ZRK<sup>78</sup>. The *given* entities are

---

<sup>76</sup> All of these aboutness relationships are without a doubt representable with the AKM since the *Aboutness* pattern allows for an aboutness relationship with any instance of the class *E1 CRM Entity*.

<sup>77</sup> Note that *general subject* is not the same as the given *general context*. As previously discussed, *general context* denotes fragments of inquiries which have not been analyzed further.

<sup>78</sup> *Zentrale Revisionskommission* (Central Auditing Commission).

particular group (“KPD”, “SED”, “ZRK”), time-span (“between 1914 and 1933”), and type of document (“reports”).

The question is a *material fact* question since it inquires about an observable relationship between a creator and documents. The provenance context is *documentation [particular group]* since the reports must have been created by the group ZRK based on a mandate. The aboutness of these reports is unknown and cannot be reasonably deduced; therefore we assign the *aboutness [any]* pattern in this case.

However, since a report does, of necessity, report on something, and a documentation activity cannot document a general subject, we can assert that the reports are either about instances of *E39 Actor* or instances of *E5 Event*.

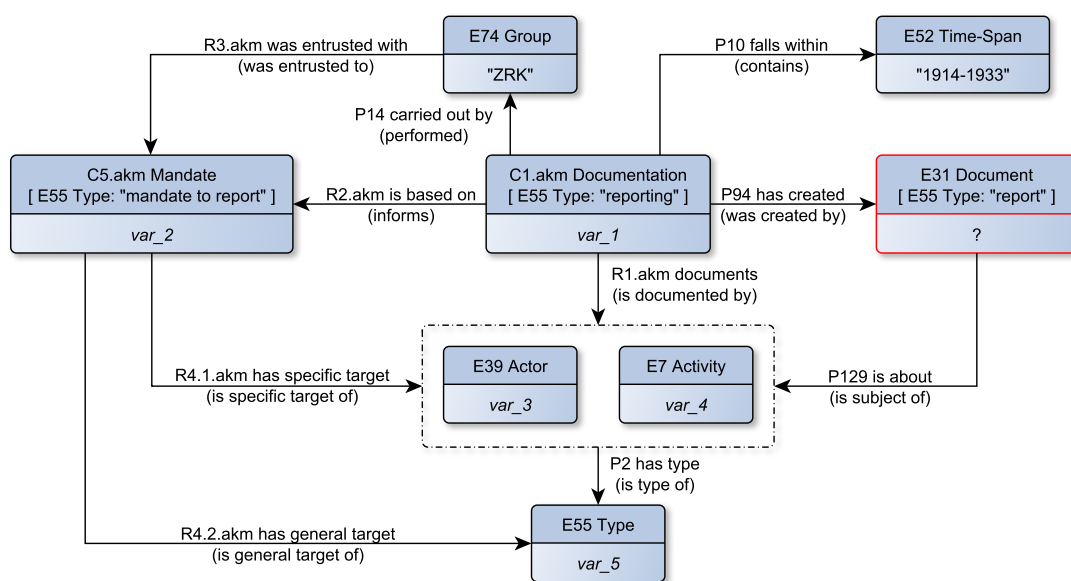


Figure 46 – Q002-01-09<sub>barch</sub>: Reports about anything.

Figure 46 shows a possible instantiation of a query pattern for Q002-01-06. A documentation activity has been carried out by the ZRK during the time-span “1914-1933” on the basis of a mandate. The activity documented actors or events and produced reports. It can be assumed that the main subjects of these reports include the documented actors or events and thus assert a relationship of *P129 is about (is subject of)* instead of the even more cautious relationship *P67 refers to (is referred to by)*. Note that an actual query would not have to include the *P129 is about (is subject of)* and *R1<sub>akm</sub> documents (is documented by)* relationship or even the *C5<sub>akm</sub> Mandate*. As mentioned at the beginning of this chapter, the example shown here represents principal instantiations showing relevant entities for a query.

Furthermore, it may also be reasonably assumed that the “reports of the ZRK” may imply reports on the ZRK’s own activities, in which case the pattern *Self-Documentation (V:2.1.3)* may



be used and may be more precise regarding the *E39 Actor* and possibly about the documented activities.

**Statistics** In the complete sample, the *Aboutness* pattern is assigned to only 41 questions (8.6%). In the BArch sample, the *Aboutness* pattern occurs 33 times which represents 9.6% of all questions in that sample. In the NAN sample, the pattern appears in eight cases representing 6.1%.

<i>Aboutness Context</i>	<i>All (n=467)</i>		<i>BArch (n=345)</i>		<i>NAN (n=131)</i>	
<b>Aboutness</b>	<b>41</b>	<b>8.6%</b>	33	9.6%	8	6.1%

Table 10 – Occurrences of the general pattern *Aboutness*.

The results show that inquiries are seldom general or unspecific regarding the topical interest of the requested archival materials. In other words, in most cases reasonable assumptions can be made about the historical reality to which a question relates.

**Summary** The *Aboutness* pattern covers the basic relationships of general reference and primary subject between an information object and some other entity. The pattern is assigned when the context of aboutness is either not specified or cannot be specified, or is irrelevant. Here, “context of aboutness” means the general or primary topic of an information object. The pattern therefore focuses on basic topical questions and inquiries where the subject relationship is too broad and thus cannot be specified. Just as in the case of the pattern *Provenance* (V:2.1.1), the pattern allows for generic and thus inclusive queries such as requests for all documents which have the GDR as their primary subject. More specific historical contexts are described in the following sub-patterns *Events* (V:2.2.2), *Plans* (V:2.2.3), *Actors* (V:2.2.4), *Things* (V:2.2.5), and *Documents* (V:2.2.6).

### 2.2.2 Events

History revolves around the intentional acts of individuals and groups and their passive or active participation in and presence at unintentional events and activities. The general pattern *Events* represents the essential constituents of the occurrence and conduct of such events and activities.

**General Pattern** The pattern *Events* represents questions about events or activities in general and things bearing information about events or activities in particular.

1. *Facts* about an event (E5) or activity (E7).
2. The identifiable but *immaterial* things (E73 *Information Object*) bearing information about an event (E5) or activity (E7).

Figure 47 shows the general *Events* pattern. The class *E5 Event* comprises phenomena which result in recognizable “changes of state in cultural, social or physical systems” (Crofts et al., 2006, 4). They have ramifications on *endurants* (IV:3.1): physical entities, such as “people, animals or things”, or conceptual entities such as “ideas, concepts, products of the imagination or common names” (Crofts et al., 2006, 4). Instances of the class *E5 Event* are not purposefully caused by actors; here the actors are passively involved or affected. Examples of instances of the class *E5 Event* are natural disasters such as earthquakes or floods, accidents, diseases and plagues, or celestial phenomena.

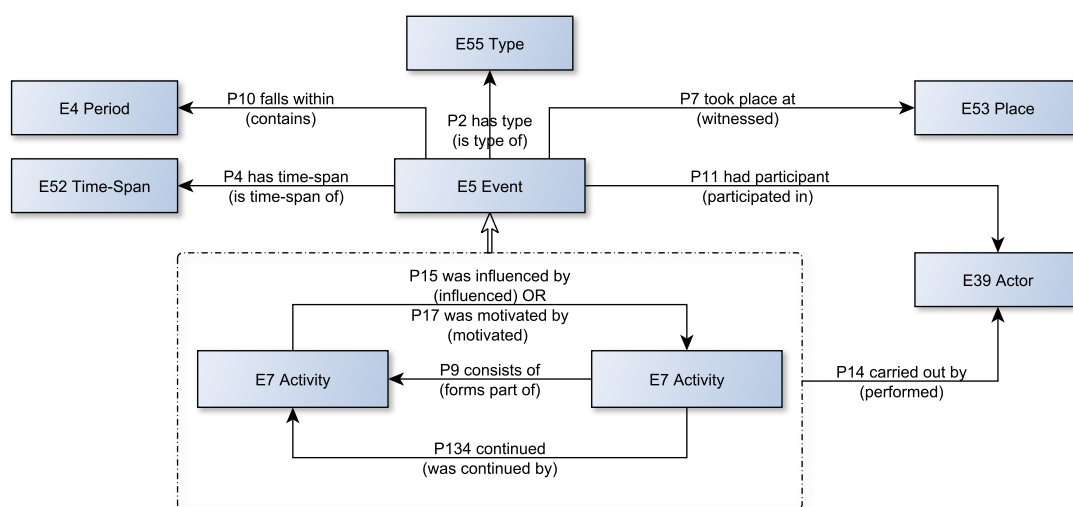


Figure 47 – General *Events* pattern.

Although users do not directly inquire about such unintentional events in their questions, such instances need to be representable in the model. For instance, users indirectly provide many examples of a type of event, namely birth and natural death events. In the CRM, these events are represented by the classes *E67 Birth* and *E69 Death* as sub-classes of *E5 Event*. Since birth and natural death events concern and directly pertain to the lives of persons, they are discussed in the context of the *Actors* pattern (V:2.2.4). For example, requests for birth or death certificates are part of the *Actors* pattern.

Instances of the class *E5 Event* can be interpreted as “non-activities” which means they happen with or without the intentional participation or immediate causation of humans. The sub-class *E7 Activity* comprises any kinds of acts which have definitely been carried out intentionally by instances of the classes *E21 Person* or *E74 Group*. Should instances of *E5 Event* that are not activities need to be distinguished, these can be referred to as “unintentional events”.

Instances of *E7 Activity* can have ramifications of any scale or magnitude within the cultural, social or physical world around the actor in question. The scale and duration of these actions may have any possible scope or duration. Examples include the Second World War, the “Day of

the Stamp”, German Reunification in 1989/1990, a parliamentary meeting of the Reichstag, or party conferences of the KPD.

Important sub-classes of *E7 Activity* are *C6<sub>akm</sub> Planning* which is discussed in the context of the *Plans* pattern (V:2.2.3), and *E86 Leaving* and *E85 Joining*, and *E66 Formation* and *E68 Dissolution*, all of which are discussed in the context of the *Actors* pattern (V:2.2.4).

The property *P11 had participant (participated in)* expresses the active or passive presence and participation of persons or groups in an event or activity.<sup>79</sup> For the purposes of this study, persons and groups only participate in actual activities as previously discussed. The only exceptions are birth and natural death events which are discussed in the context of the *Actors* pattern (V:2.2.4).

Examples for the active or passive participation or presence of an actor in an activity include the visit (P11) of a delegate to a party conference, a soldier’s participation (P11) in a battle, or the act of receiving (P11) a medal.

The property *P14 carried out by (performed)* expresses the active and *causal* participation in or performance of an activity by a person or group. The aspect of causal participation or performance must be construed broadly: For example, creation activities, as discussed in the context of the *Provenance* pattern (V:2.1.1), are carried out (P14) by the person formulating a personal letter or a group collectively taking (P14) the minutes of a meeting. In a more abstract sense, J.F. Kennedy carried out (P14) his visit to Berlin in 1968 while photographers only participated in (P11) this activity.

Furthermore, the *Events* pattern focusses on the event or activity itself. Questions which are about the actor his or herself are part of the *Actors* pattern (V:2.2.4).

Events and activities happen (*P7 took place at (witnessed)*) at particular places (E53) and occur (*P4 has time-span (is time-span of)*) during particular time-spans (E52) or fall within (*P10 falls within in (contains)*) particular periods (E4). For example, the writing of the letter occurred during the summer of 1946 (E52), or the parliamentary meetings of the Reichstag took place in the Reichstag building (E53) in Berlin (E53) during the Weimar Republic (E4).

Questions often demand documents about activities which occurred in a particular *context* of another activity. For this reason, connecting activities is important and also allows for the discovery of additional documents which may be connected to intermediate activities.

The *Events* pattern covers three principal relationships between activities:

1. Forms of *part-whole* relationships between activities are expressed by the property *P9 consists of (forms part of)*. Each activity which forms a part of a larger activity is a *constitutive part* of the logical whole. For example, the policy of socializations (E7) of private businesses during the early years of the GDR were part of the general process of “Stalinisierung” (E7) in Central Europe.

---

<sup>79</sup> Note that the property *P11 had participant (participated in)* also pertains to the class *E7 Activity* since this class is a sub-class of *E5 Event*.

2. Forms of *sequences* of activities where one activity is the *intentional continuation* of another are expressed by the property *P134 continued (was continued by)*. For example, the parliamentary deliberations on a new law may continue over several different activities such as various committee meetings and plenary assemblies.
3. Forms of general *causal influence* between two or more activities are expressed by the property *P15 was influenced by (influenced)*. For example, a regular meeting of town officials is influenced by the bombing of their city. If an activity happens only *in reaction to* another, then the property *P17 was motivated by (motivated)* can be used; for example, a statement (E7) is given in reaction to (P17) public critique of the enactment (E7) of a new law.

Part-whole relationships permit the modelling of situations where activities happen in the specific context; that is, as part of a larger activity. These relationships are important for requests for documents about activities which occurred as part of another.

Sequences and general causal relationships allow for the modelling of simple *before* and *after* situations. For inquiries requesting documents about activities which either happened in reaction to or in continuation of a previous activity, such connections are essential.

As a specialization of the *Events* pattern, the pattern *Plans* describes an even more explicit and intentional causation between two or more activities: the general and specific purpose of a planning activity and the planned activity. Since plans are pivotal in this context, these situations are part of the *Plans* pattern.

In contrast to the *Actors* pattern (V:2.2.4), questions pertaining to the *Events* pattern primarily focus on the event or activity. The actor is not the main interest of the question. Questions pertaining to the *Actors* pattern may also inquire about activities; however, these recognizably place their interest on the actor and the activities carried out by him, her or them.

In terms of query instantiations, questions pertaining to the *Events* pattern will ask for documents about events or activities, while questions pertaining to the *Actors* pattern will ask for documents directly about an actor and about activities carried out by him, her or they.

The following examples demonstrate various instantiations of the *Events* pattern and discuss different aspects of the interpretation of questions.

#### **Example - “Photographs of a particular activity”**

Q031-01-03<sub>barch</sub> “Pictures of J.F. Kennedy’s visit to Berlin, June 1963.”

The general question type is *resource discovery <specific type>*. The *wanted* entity is *<activity>*. The relevant *given* entities are *type of document* (“pictures”), *particular activity* (“J.F. Kennedy’s visit to Berlin, June 1963”), *particular person* (“J.F. Kennedy”), *particular place* (“Berlin”), and *time-span* (“June 1963”).

The question is a *material fact* question since it inquires about an observable relationship between pictures and an activity. The *provenance context* is *documentation [actor]* since the

pictures of the visit of J.F. Kennedy to Berlin must have been created as a direct observation of the activity. However, the creator is unknown. The *aboutness context* is *events [particular activity]*.

Figure 48 shows the most simple variant of the example. Here, a query would ask for any visual item, such as private photographs, depictions in newspaper articles, or even drawings, about the visit of J.F. Kennedy to Berlin in June 1963.

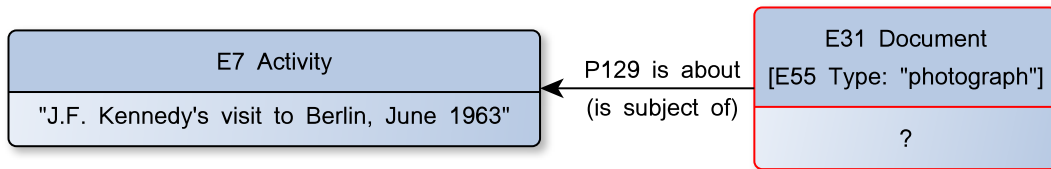


Figure 48 – Q031-01-03a<sub>barch</sub>: Photographs of a particular activity.

This example resembles a plain subject access query. The query will succeed if the activity “J.F. Kennedy’s visit to Berlin, June 1963” (E7) has been identified and explicitly indexed, as well as related to appropriate depictions (E36) of the activity. This is unlikely to be the case in most archival information systems. Alternatively, a keyword search might succeed if appropriate text elements, such as “Kennedy”, “visit to Berlin”, and “pictures”, appear close enough to each other in the metadata; that is, in the description of the archival unit which contains relevant depictions.

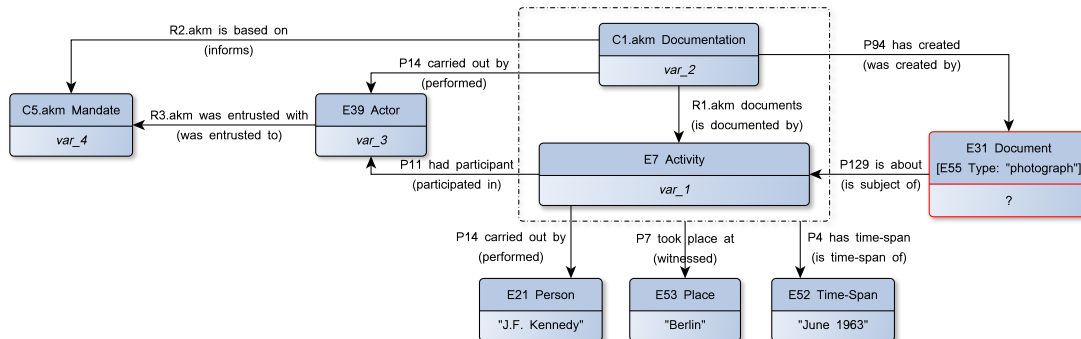


Figure 49 – Q031-01-03b<sub>barch</sub>: Photographs of a particular activity as the result of a documentation activity.

Figure 49 is meant to demonstrate the pattern-based approach to a relevant query based on the known contextual information. In this figure, the known contextual information and the known implicit knowledge from the ontology are explicitly shown. The previously assumed particular activity “J.F. Kennedy’s visit to Berlin, June 1963” (E7), however, is not given but identified by contextual information.

The activity is identified by the participation of the person “J.F. Kennedy” (E21), by the place where the activity happened, “Berlin” (E53), and by the time-span during which the activity took place, the duration of “June 1963” (E52). The assumed documentation activity, from which the pictures (E36) of the visit must have resulted, is also further qualified by the same place and time-span.

The documentation activity must have been carried out by an actor – a group or person – who at the same time *participated* in the visit of J.F. Kennedy. This actor might have taken photographs based on a mandate, for example, a newspaper reporter or spies from the Soviet side.

Any of this information will help to discover and identify relevant visual material documenting the activity in question. This example demonstrates how the patterns and the AKM indicate which kinds of information are relevant to the user and which kinds of information should be made explicit.

### Example - “Documents about a particular activity”

Q023-12-02<sub>barch</sub> “Do you have documents about the topic ‘Bombing of Kassel 1943?’”

The general question type is resource *discovery* <*material-finding*>. The *wanted* entity is <*activity.role*>. The relevant *given* entities are *particular activity* (“bombing [of Kassel 1943]”). The *supplementary* information includes *time-span* (“1943”), and *particular place* (“Kassel”).

This is a *material fact* question since it inquires about an observable subject relationship between a document and an activity. The *provenance context* is *provenance [actor]* since no indication is given as to who might have created a document about the bombing of Kassel in 1943. The *aboutness context* is *events [particular activity]* since “bombing of Kassel 1943” designates a particular activity.

The question is unspecific and general. The first question is whether “Bombing of Kassel 1943” refers to all bombings of Kassel which might have happened during the year 1943, or a one particular bombing. Since “topic” indicates an interest in any activity of the type “bombing”, a type of activity is assumed.

Two principle interpretations are possible: (1) The first interpretation includes any documents resulting from a direct or indirect observation of the activity “Bombing of Kassel 1943”, such as diaries, statements of eye witnesses, newspaper articles, or official reports on damages etc. Direct or indirect observations only entail that the documents are directly about or directly refer to the activity “Bombing of Kassel 1943”. (2) The second principal interpretation would understand the request to be for any documents about activities “influenced” by the “Bombing of Kassel 1943”, such as minutes of meetings discussing the events or speeches in reaction to the bombings.

Figure 50 shows the first principal interpretation. The diagram is similar to that of Figure 49 in that the given contextual information is explicitly represented (“Kassel” (E53) and “1943” (E52)), and a documentation activity (C1<sub>akm</sub>) is shown. In this example, the actor and the

mandate have been omitted. Of course, the most simple instantiation would be any document (E31) about (P129) the “Bombing of Kassel 1943” (E7); however, the same problems would apply as discussed in the context of the previous example and Figure 48.

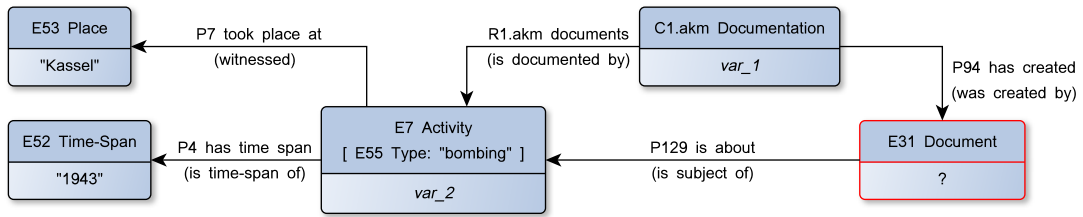


Figure 50 – Q023-12-02a<sub>barch</sub>: Documents about a particular activity.

An example of such a document would be any newspaper article (E31) about (P129) the “Bombing of Kassel 1943” or photographs taken of the events (E31). These documents may have been created during but also at any time after the event.

Figure 51 shows the second principal interpretation. This instantiation covers any documents about (P129) any activities (E7) which happened after the “Bombing of Kassel 1943” but which are related (P15) to it.

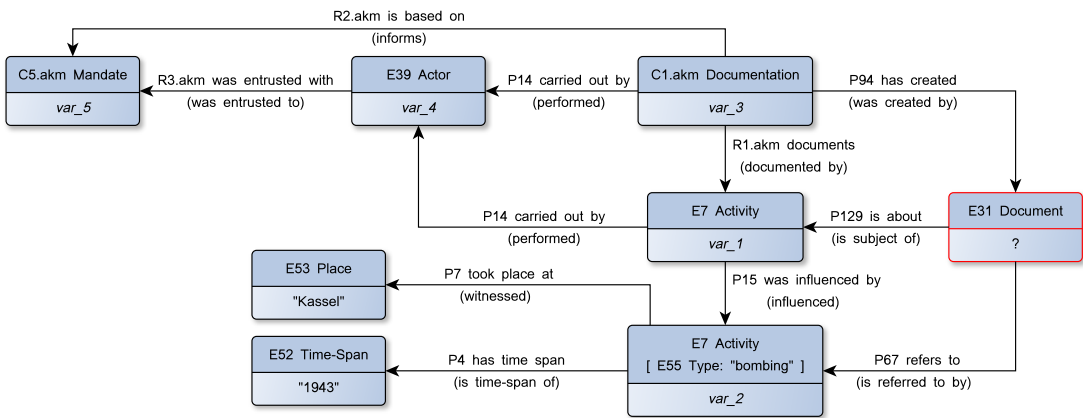


Figure 51 – Q023-12-02b<sub>barch</sub>: Documents about a particular activity.

Again, the most simple instantiation would be any documents (E31) about (P129) any activity (E7) influenced by (P15) the events of the “Bombing of Kassel 1943” (E7); for example, the

removal of rubble and ruins (E7) would be an activity motivated by (P17) the “Bombing of Kassel 1943” (E7). These activities may have been documented in diaries, which would probably also refer directly to the bombings, or newspaper articles. The same instantiation would be possible with  $C2_{akm}$  *Self-Documentation* which would cover any meetings (E7) influenced by (P15) the “Bombing of Kassel 1943” (E7).

### Example - “Documents about a named activity”

Q059-01-12<sub>barch</sub> “I am interested in documents of the philatelists in the Kulturbund of the SBZ and GDR which are thematically about the ‘Day of the Stamp’ and which are limited to the time approximately until the end of October 1949.”

The general question type is *resource discovery* <material-finding>. The *wanted* entity is <activity>. The relevant *given* entities are *particular group* (“SBZ”, “GDR”, “Philatelists in the Kulturbund”), *particular activity* (“Day of the Stamp”), and *time-span* (“ca. until the end of October 1949”). This is a *material fact* question since it inquires about an observable subject relationship between documents and an activity. The *provenance context* is *provenance* [*particular group*] since the question requests documents either kept or created by a particular group. The *aboutness context* is *events* [*particular activity*].

Figure 52 shows the exemplary query pattern. Regarding the provenance context, particular queries may ask for things which have been either created (P108), produced (P94), or kept (P49) by the particular group “philatelists in the Kulturbund”. The actor subsumes not only the particular group “philatelists in the Kulturbund” but also by any person or group as a member of this particular group.

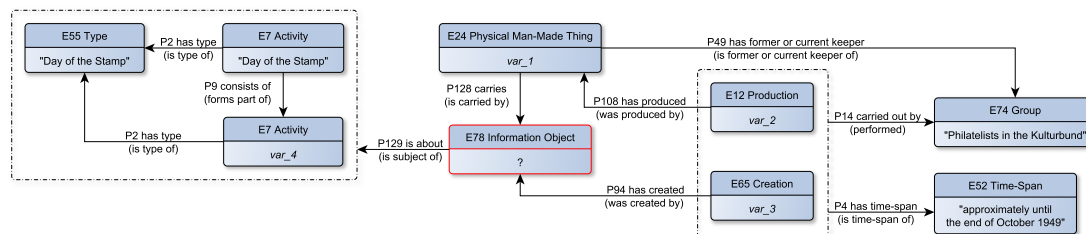


Figure 52 – Q059-01-12<sub>barch</sub>: Documents about a named activity.

This basic pattern is known and explained in detail in the *Provenance* pattern (V:2.1.1). Additional important given information includes the time-span “approximately until the end of October 1949”, during which the production and creation activities took place.

Regarding the aboutness context, the target of any query consists of information objects about the particular activity “Day of the Stamp”. This particular activity can be represented in various ways:



1. information objects about the particular activity “Day of the Stamp”, if one particular “Day of the Stamp” is assumed, this would be to assert the latter, subsuming all the “Day of the Stamp” activities which have ever occurred,
2. information objects about any activity which was part of the particular activity “Day of the Stamp”; for example, various activities organized in the context of the “Day of the Stamp”, including those within the general context of another activity,
3. or information objects about activities which generally occurred in the context of a principal activity of the type “Day of the Stamp”, such as the particular activities each year in the context of the annual “Day of the Stamp”.

This example should demonstrate that varying representations of specific activities are possible and depend on the specific activity to be modelled. The third option would probably be the most relevant in the current context; of course, this always depends on the data pertaining to the specific instance that is to be represented with the AKM. The AKM, however, is flexible enough to accommodate most reasonable variations. The *Plans* pattern, which is introduced next, would represent another possible query specifically about the planning the “Day of the Stamp”.

**Statistics** In the complete sample, the *Events* pattern has been assigned to 109 questions representing 23% of all inquiries. Of these, 70 questions belong to the BArch sample amounting to 20% of all inquiries in that sample, while the remaining 39 questions represent 30% of the NAN sample. The general pattern occurs only once in the BArch sample with the unspecific primary entity of interest *any*. Most important in the whole sample and in the context of the *Events* pattern are *particular activities* with 75 occurrences (69%), of which 46 (66%) belong to the BArch sample and 29 (74%) to the NAN sample. Second comes *type of activities* amounting to 31 (28%) in the whole sample of which 23 (33%) belong to the BArch sample and 8 (21%) to the NAN sample. Only two questions (5.1%) referred to a *particular event* in the NAN sample, while *type of event* did not occur at all.

<i>Aboutness Context</i>	<i>All (n=467)</i>		<i>BArch (n=345)</i>		<i>NAN (n=131)</i>	
<b>Events</b>	<b>109</b>	<b>23%</b>	70	20%	39	30%
- Events [ any ]	1	0.9%	1	1.4%	0	-
- Events [ particular activity ]	75	69%	46	66%	29	74%
- Events [ type of activity ]	31	28%	23	33%	8	21%
- Events [ particular event ]	2	1.8%	0	-	2	5.1%
- Events [ type of event ]	0	-	0	-	0	-

*Table 11* – Occurrences of the general pattern *Events*. Note that percentages for the primary entities of interest are per current general pattern.

The figures indicate the importance of questions generally pertaining to the context of events and activities, specifically to particular activities, and to a lesser degree types of activities.

**Summary** The *Events* pattern distinguishes between unintentional events and intentional acts carried out by actors such as persons or groups. Both can be described in their temporal and spatial contexts, where and when an activity happened and who participated. Complex activities are described by the relevant and fundamental relationships of influence, sequence, and part-whole. The pattern thus allows us to model basic activity-centred history. The pattern covers questions such as: What happened and how did an activity or event occur? What happened to persons or groups? Who acted and how? When and where did these activities and events happen?

The pattern focuses on requests for archival material which may provide evidence for the execution or existence of activities, or factual and descriptive accounts on the act structure of activities or events; that is, what happened and how.

The pattern is therefore assigned when a question inquires after information objects about historical activities, or documents which describe or otherwise give information or evidence on these activities.

While the *Events* pattern does allow us to ascertain a motivational or influential relationship between activities, it does not, however, represent a more detailed historical context of activities. Inquiries about information objects that would provide information and evidence on reasons for activities, who enacted and desired them to be carried out, and whether or not activities adhered to the original plan, lie outside the scope of this pattern. The planning and wanting of activities, the will and the plan, are covered by the sub-pattern *Plans*, described next.

### 2.2.3 Plans

The *Events* pattern describes unintentional events and intentional acts carried out by people. These activities typically follow specific plans, and persons or groups express the will to do something, to enact or not to enact a plan.

**General Pattern** The *Plans* pattern represents questions about the planning and execution of plans in general, and things *about* planning, plans, will, and the execution of plans. The pattern *Plans* covers questions relating to five cases:

1. The planning ( $C6_{akm}$ ) of activities (E7) such as a meeting preparing the agenda for the next party conference.
2. The creation of plans ( $C4_{akm}$ ) such as initial drafts of architectural drawings.
3. The discussion, decision and abandonment of plans ( $C4_{akm}$ ) such as deliberations in parliamentary committees about a new law.

4. The identifiable but immaterial expression of a will ( $C3_{akm}$ ) to execute or realize a plan ( $C4_{akm}$ ), such as a parliamentary resolution to enact a new law.
5. Activities (E7) which follow a specific or general plan ( $C4_{akm}$ ), such as the enforcement of a law.

Figure 53 shows the principle *Plans* pattern.

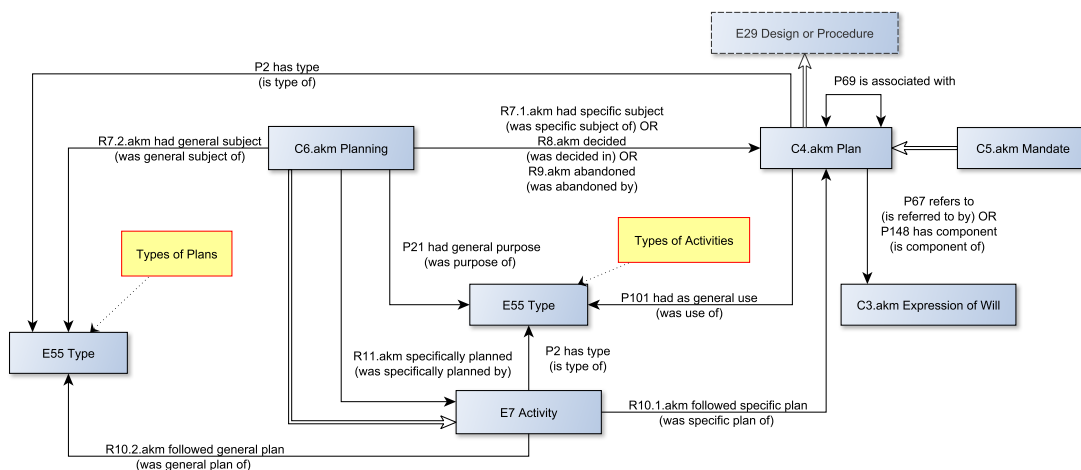


Figure 53 – General Plans pattern

The class  $C4_{akm}$  Plan is a sub-class of  $E29$  Design or Procedure and represents the intellectual structure of a plan. This intellectual structure, or plan structure, constitutes a schema for the execution of activities (E7). Three principal types of plan structures can be distinguished:

1. descriptions of *appearance*, for example a plan of a building or a topographical map;
2. descriptions of *procedure*, for example a legal regulation, law or mandate;
3. descriptions for achieving a condition or future state of affairs, for example a political agenda or policy.

Particular examples of instances of the class  $C4_{akm}$  Plan include, for instance, “the social policy of the British government in the 19th century”, “the plan of the SED to solve the housing problem in the GDR”, or “the ground floor plan of the Reichstag”. Accordingly, plans may have (P2) a general type (E55), such as “social policy”, “law” or “building plan”.

Instances of the class activity (E7) may follow or execute a particular plan. For example, the building of new housing in East Berlin followed “the plan of the SED to solve the housing problem in the GDR”. This relationship is expressed by the property  $R10.2_{akm}$  followed specific plan (was specific plan of).

Accordingly, an activity may follow a general type of plan; for example, the particular activity of building new housing in East Berlin followed a plan of the type “social policy”. This relationship is expressed by the property *R10.1<sub>akm</sub> followed general plan (was general plan of)*.

Particular activities may also have a general type (E55); for example, the particular activity “the trip to France of the FDGB” (E7) has the principle activity type “travel” (E55). Consequently, a particular plan may have a type of activity as its general usage scenario. This relationship is expressed by the property *P101 had as general use (was use of)*. For example, the “plan of the FDGB to travel to France” (*C4<sub>akm</sub>*) has as its general usage (P101) the planning of activities of the type “travel” (E55).

Continuing with this example, the type of particular “plan of the FDGB to travel to France” (*C4<sub>akm</sub>*) would be “travel plan” (E55). Then, the particular activity “the trip to France of the FDGB” (E7) would have followed (*R10.1<sub>akm</sub>*) the plan type “travel plan” (E55) and executed (*R10.2<sub>akm</sub>*) the particular “plan of the FDGB to travel to France” (*C4<sub>akm</sub>*).

These properties allow for detailed descriptions as to how particular plans and types of plans (*C4<sub>akm</sub>*) and activities (E7) relate to each other. In this regard, the *Plans* pattern extends and specializes the *Events* pattern (V:2.2.2) with the notion of a purposeful and planned execution of activities.

The class *C6<sub>akm</sub> Planning* is a sub-class of *E7 Activity*. Instances of the class *C6<sub>akm</sub> Planning* represent activities with two general notions:

1. planning activities which plan and prepare other activities, for example: a planning meeting of the FDGB prepares travel to France
2. planning activities which discuss plans

Instances of the class *C6<sub>akm</sub> Planning* may plan or prepare a particular activity (E7) that is either known or assumed to have taken place. This relationship is expressed by the new property *R11<sub>akm</sub> specifically planned (was specifically planned by)*. For example, the meeting (*C6<sub>akm</sub>*) of the FDGB prepared (*R11<sub>akm</sub>*) “the trip to France of the FDGB” (E7). The new property *R11<sub>akm</sub>* is a sub-property of the property *P20 had specific purpose (was purpose of)* and extends its semantics: While P20 expresses that the activity succeeded in achieving its aim, *R11<sub>akm</sub>* is asserted when it is known that an activity occurred, independently from the question as to whether the activity succeeded in achieving its specific goals. Activities conducted only in part are regarded as having been carried out. This extension appears necessary since in the current historical context it is not possible to ascertain whether or not activities succeeded in achieving their goals. It is, however, important to be able to assert a categorical link between a planning activity and a activity which has been planned.

The general purpose of a planning activity may be expressed by the property *P21 had general purpose (was purpose of)*. For example, the planning meeting (*C6<sub>akm</sub>*) of the FDGB had as its general purpose (P21) the preparation of activities (E7) of the type “travel” (E55).

Planning activities may not only prepare particular activities or types of activities but, more importantly, they also discuss the plans guiding the execution of the planned activities.

Particular plans may have been the subject of a planning activity. This relationship is expressed by the property *R7.1<sub>akm</sub> had specific subject (was specific subject of)*. For example, during a series of planning meetings (*C6<sub>akm</sub>*) the FDGB discussed and subsequently devised (*R7.1<sub>akm</sub>*) the particular “plan of the FDGB to travel to France” (*C4<sub>akm</sub>*).

Accordingly, a series of planning meetings can be said to have discussed general types of plans. This relationship is expressed by the property *R7.2<sub>akm</sub> had general subject (was general subject of)*. For example, the FDGB regularly conducted meetings (*C6<sub>akm</sub>*) which were, among other things, dedicated to (*R7.2<sub>akm</sub>*) the discussion and preparation of “travel plans” (*E55*), as a specific type of plan, for the various youth groups in the organization.

The possible relationships between the three entities of planning activities, plans, and activities executing plans can be summarized as three general groups of relationships:

1. general and specific aboutness, which expresses the topical relationship between planning and the plan itself
2. general and specific purpose, which expresses preparatory relationships between planning and the activity
3. general and specific usage, which expresses the relationship between the plan and the execution of the activity.

The following parts discuss the creation of plans (*C4<sub>akm</sub>*) and of expressions of will (*C3<sub>akm</sub>*), including the notion of deciding upon and abandoning plans.

Figure 54 shows how the class *E65 Creation* and its sub-classes *C1<sub>akm</sub> Documentation* and *C2<sub>akm</sub> Self-Documentation* relate to the classes *C6<sub>akm</sub> Planning*, *C4<sub>akm</sub> Plan*, and *C3<sub>akm</sub> Expression of Will*.

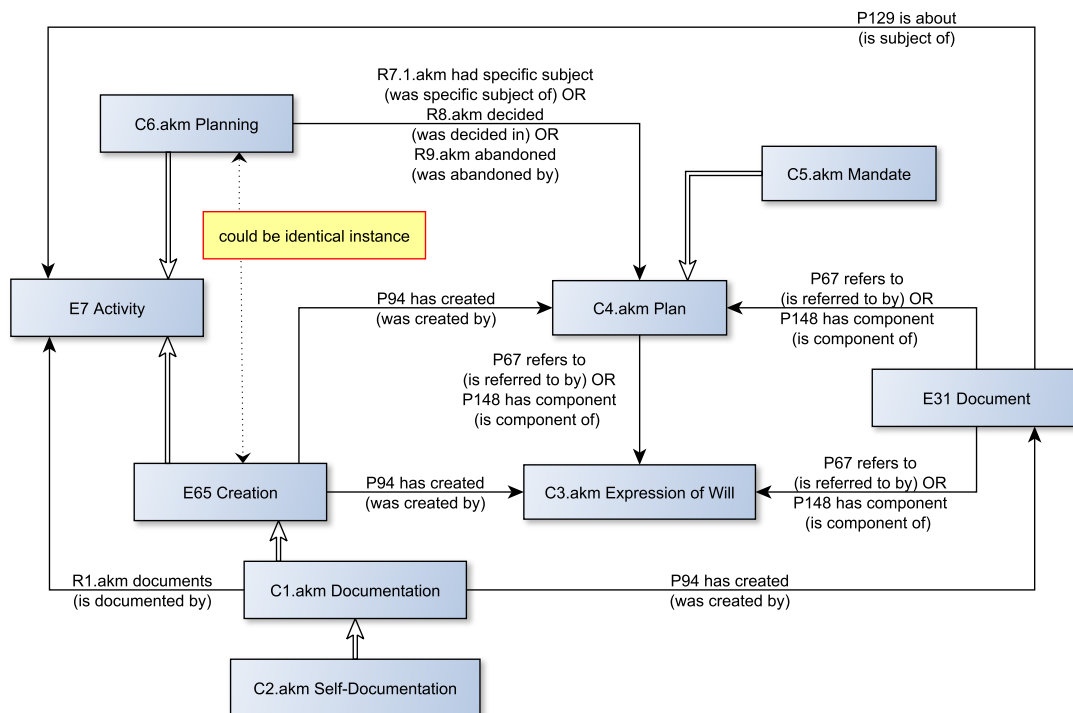


Figure 54 – General Plans pattern: Planning, plans, and expressions of will.

Planning activities ( $C6_{akm}$ ) may not only discuss an existing plan ( $C4_{akm}$ ) but actually create a new one ( $C4_{akm}$ ), or modify or supplement an existing plan, thus creating a new, recognizable version. For example, a parliamentary committee may create the first draft of a new law ( $C4_{akm}$ ) and, in subsequent meetings<sup>80</sup>, the committee may create new versions of the law ( $C4_{akm}$ ).

If an instance of the class  $C6_{akm}$  Planning creates an initial or new version of a plan, then it is also an instance of the class  $E65$  Creation. Then the planning activity not only discusses ( $R7.1_{akm}$ ) a particular plan but also creates ( $P94$ ) intellectual contents of a plan.

Furthermore, planning activities ( $C6_{akm}$ ) may decide either to enact or to abandon a plan ( $C4_{akm}$ ). The former notion is expressed by the new property  $R8_{akm}$  decided (*was decided in*) and the latter is expressed by the new property  $R9_{akm}$  abandoned (*was abandoned by*). Both properties are sub-properties of  $R7.1_{akm}$  had specific subject (*was specific subject of*) since they are more specific regarding the relationship between the planning activity and the plan.

In both cases, the planning activity ( $C6_{akm}$ ) is at the same time an instance of the class  $E65$  Creation, because an instance of the class  $C3_{akm}$  Expression of Will is created. The psychological entity *will* has been discussed in the context of the “Ontological Core Framework” (V:1.4). The will to do something – for example, to execute or to abandon a plan – becomes evident through

<sup>80</sup> The representation of sequences of activities is described in the Events pattern (V:2.2.2).

an instance of the class  $C3_{akm}$  *Expression of Will*.

For example, the passing of a law ( $C4_{akm}$ ) by parliamentary vote is the expression of the will of the parliament to enact the law. In this example, the session of the parliament is an instance of the class  $C6_{akm}$  *Planning*, which enacts ( $R8_{akm}$ ) the law by vote and, at the same time, the session is an instance of the class  $E65$  *Creation* because the vote creates (P94) a formal resolution ( $C3_{akm}$ ) as an expression of the will of the parliament to do so. This resolution ( $C3_{akm}$ ) constitutes evidence of the will of the parliamentary body.

Laws and legislative proposals are also instructive examples of why the plan and the expression of will are distinguished. While parliamentary committees prepare the legislative proposal, only the parliamentary body performs an act of will by either passing or rejecting the proposal. This act of parliament ( $C3_{akm}$ ) is legislatively separate from the preparatory deliberations ( $C6_{akm}$ ) of the proposal ( $C4_{akm}$ ).

While the enactment ( $R8_{akm}$ ) of a plan also marks the *beginning* of a will, for example the will of the parliament to enact a law, the *end* of a will is marked by the abandonment ( $R9_{akm}$ ) of a plan; for example the annulment of the same law by parliament, or, as very different kind of planning activity, the annulment ( $R9_{akm}$ ) of a marriage as the end of the will to be a legally married couple.

It is a quite different question, of course, whether the plan has actually been enforced or executed, whether or not the activity which follows the plan does so in compliance, or whether a plan has even been abandoned. Here, a range of historical documents can be pertinent, such as public statements contradicting public actions.

A will can also be expressed through activities other than planning activities. For example, public statements or proclamations are activities ( $E65$ ) which create an expression of will. Relevant historical issues include the question as to whether perhaps a particular will was expressed despite the fact that it never existed, or whether the execution of a plan shows that a will must have existed despite their being no specific expression of that will.

Plans and expressions of will are information objects (V:2.1.1), and may therefore refer to each other (P67) or be part of (P148) another information object such as documents (E31).

Figure 54 shows the principal referral and component relationships which may exist between plans, expressions of will, and documents. For example, the plan will typically be the component (P148) of a document (E31), such as minutes, which have resulted from the self-documentation ( $C2_{akm}$ ) of a planning meeting ( $C6_{akm}$  /  $E65$ ).

In principle, the resolution ( $C3_{akm}$ ) may be referred to by (P67) or be part of (P148) various information objects (E78) such as the minutes (E31) of the parliamentary session as a result of self-documentation ( $C2_{akm}$ ), or it may be mentioned in the official publication (E78) of the law.

Expressions of will may also be documented and indirectly referred to, for example in newspaper articles or personal accounts of the event.

Plans may, of course, depend on each other, appear in sequence, or form part of another plan. If necessary, such relationships may be modelled with the symmetric property  $P69$  is

*associated with*, which expresses an unspecified relationship between two instances of *E29 Design or Procedure*.

The class *C5<sub>akm</sub> Mandate* is a sub-class of the class *C4<sub>akm</sub> Plan*. Mandates have been discussed in detail in the context of the *Documentation* pattern (V:2.1.2). All of the above also pertains to mandates: They are created, discussed, decided upon and abandoned, as well as 'wanted' through planning and creation activities.

### Example - "Planning an activity"

Context                    *"In 1980, a delegation of the FDGB lead by Harry Tisch laid down a wreath of flowers in Oradour. The visit was part of a trip of the FDGB to France (...). At this time, Tisch was also a member of the Politbüro of the ZK of the SED."*

Q005-08-01<sub>barch</sub>        *"Where can I find documents about the planning [of this trip] (...)?"*

The general question type is *resource discovery <material-finding>*. The *wanted* entity is *<activity>*. The relevant *given* entities are *particular person* ("Harry Tisch"), *particular place* ("Oradour", "France"), *particular activity* ("trip [of the FDGB to France]"), *particular group* ("FDGB", "Politbüro [of the ZK of the SED]", "SED"), *time-span* ("1980"), and *type of activity* ("planning").

This is a *material fact* question. The *provenance context* is *self-doc [actor]* since it remains open whether a group or person created documents pertaining to the planning of the trip. The *aboutness context* is *plans [particular activity]* since the question inquires about documents related to the planning of a particular activity.

The main interpretation of the question is that documents (E31) which may give information about the planning of the particular activity "trip [of the FDGB to France]" (E7) must have been the result of any kind of self-documentation of an actor who documented his or her planning activities.

As discussed in the *Events* pattern (V:2.2.2), the particular activity "trip of the FDGB to France" (E7) is indirectly identified by a variety of given contextual information, such as the place of the activity ("France"), the time-span ("1980"), the principle type of activity ("trip"), and various participants. This contextual information is not explicitly represented in Figure 55 but a distinct and particular activity "trip of the FDGB to France" is assumed.<sup>81</sup>

---

<sup>81</sup> In principle, the trip could also be easily modelled as a sequence of activities using the means provided by the *Events* pattern (V:2.2.2). This, however, depends on the richness of the particular knowledge base and the desired level of detail of the instance data.



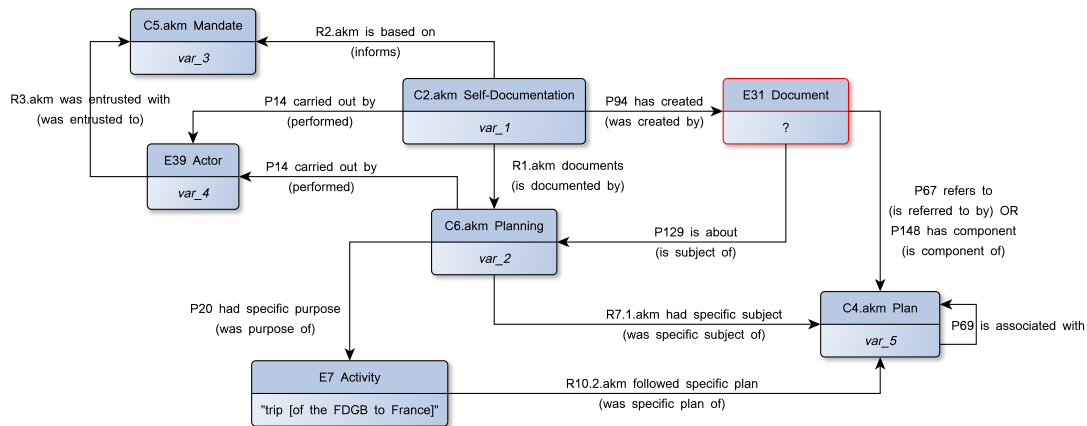


Figure 55 – Q005-08-01<sub>barch</sub>: Documents about planning activities for a specific event.

The question demands several queries in order to retrieve a set of relevant documents about the planning of the trip of the FDGB to France. These queries are summarized in Figure 55. Primarily three different queries regarding particular documents can be distinguished:

1. Documents (E31) making statements (P67) about (P129) a particular plan (C4<sub>akm</sub>) which has been
  - (a) either executed (R10.2<sub>akm</sub>) during the particular activity “trip [of the FDGB to France]” (E7)
  - (b) or discussed (R7.1<sub>akm</sub>) during a planning meeting (C6<sub>akm</sub>) which had the specific purpose (P20) of planning the particular activity “trip [of the FDGB to France]” (E7).
2. Documents (E31) which include (P148) such a particular plan (C4<sub>akm</sub>).
3. Documents (E31) which are about (P129) such a planning activity (C6<sub>akm</sub>).

In all cases, the documents (E31) would have been most likely the result of a self-documentation (C2<sub>akm</sub>). Since the actor who carried out the planning and self-documentation activity is unknown – and this is an important point – an *E39 Actor* is given in this query pattern. This means that a particular query could either leave the *E39 Actor* unspecified and thus include any instances of *E21 Person* or *E74 Group*, or it may in turn instantiate the class *E39 Actor* with the particular groups “FDGB”, “Politbüro [of the ZK of the SED]”, “SED”, or the particular person “Harry Tisch”.

In the context of “1a” and “1b”, of course, a query could also directly ask for a particular plan as a distinct information object. Finally, several related plans may exist which influenced the trip to France or were discussed during planning meetings. Queries may therefore also seek associated (P69) plans.

The query pattern in Figure 55 permits several entry points for queries. The most relevant and instructive have been discussed. Additional possible indirections include instances where the decision has been taken during a planning meeting to execute (R8<sub>akm</sub>) or to abandon (R9<sub>akm</sub>) the plan. In both cases, an expression of will (C3<sub>akm</sub>) would also have been created.<sup>82</sup>

Furthermore, beginning with the mandate (C5<sub>akm</sub>), it would be possible to inquire as to who had the mandate to plan activities of the type “travel”. All indirections, however, are within the general query pattern as described in Figure 53 and 54.

The next example originates from the same inquiry and will demonstrate how apparently very different kinds of questions can be answered with the same basic query patterns.

### Example - “Discussions”

Context                    *“In 1980, a delegation of the FDGB lead by Harry Tisch laid down a wreath of flowers in Oradour. The visit was part of a trip to France of the FDGB (...). At this time, Tisch was also a member of the Politbüro of the ZK of the SED.”*

Q005-08-03<sub>barch</sub> *“In your opinion, was such a trip ever discussed or, at least, approved by the ZK?”*

The general question type is *fact-finding*. The *wanted entity* is *<group.activity>*. The relevant *given entities* are *particular person* (“Harry Tisch”), *particular place* (“Oradour”, “France”), *particular activity* (“trip [of the FDGB to France]”), *particular group* (“FDGB”, “Politbüro [of the ZK of the SED]”, “SED”), *time-span* (“1980”), and *type of activity* (“planning”).

This is a *material fact* question. The *provenance context* is *self-doc [particular group]* since potentially relevant documents must have been created as a result of self-documentation by the particular group “ZK”. The *aboutness context* is *plans [particular activity]* since the question inquires about documents related to the planning of a particular activity. The question asks whether the particular group “ZK”<sup>83</sup> had discussed or approved the particular activity “trip to France of the FDGB” (E7).

First of all, it is important to remember that the patterns are about the general and generic relationships between certain entities and not about the many specific qualities of these connections: it is not relevant whether the relationship between a document and an activity is one of “discussion” or “approval”. Rather, on the most generic semantic level, it is a relationship of “aboutness”. It is the genuine task of the researcher to read and interpret the documents in order to find out about the qualitative aspect: whether the ZK did in fact “discuss” or, more to the point, “approve” something. The pattern is a means for the researcher to discover potentially relevant facts and documents.

Since this is a factual question, an answer would retrieve material facts showing that the particular group “ZK” or a member of this group, either a person or another sub-group, carried

---

<sup>82</sup> An example for expression of will is given on page 198.

<sup>83</sup> “ZK” is the abbreviation for the *Zentralkomitee* (central committee) of the *Sozialistische Einheitspartei Deutschland* (SED).

out an activity which either had the specific purpose of planning the particular activity “trip to France of the FDGB” (E7) or discussed a plan on which the particular activity “trip to France of the FDGB” (E7) or any related activity was based. This factual context is shown in Figure 56. In the interests of clarity, a particular “plan [for the trip to France of the FDGB]” (C4<sub>akm</sub>) is assumed for this exemplary query pattern. Of course, plans associated with the specific “plan [for the trip of the FDGB to France]” are also potentially relevant.

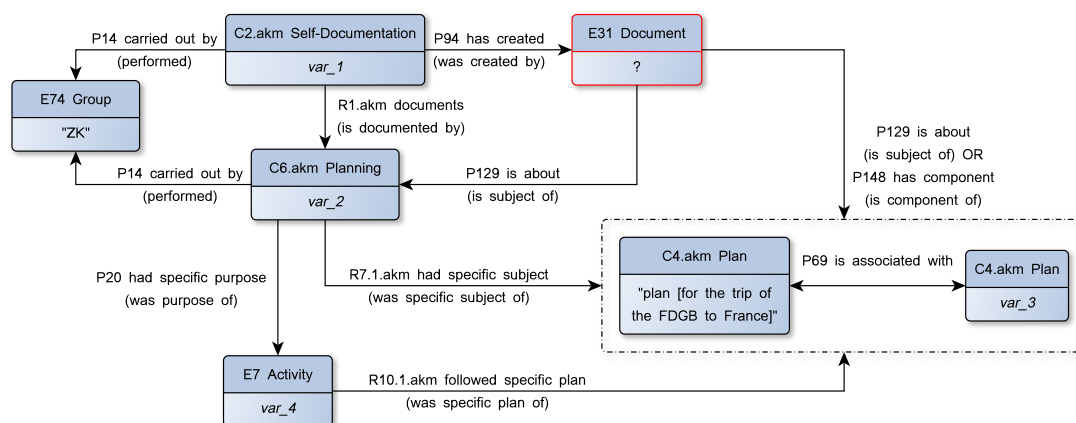


Figure 56 – Q005-08-03<sub>barch</sub>: Documents attesting discussion or approval of a plan.

For factual questions, this study assumes that information objects are indirectly desired by the user in order to establish evidence for facts. The same principal documents as discussed in the previous example on page 190 would be relevant with regard to information objects. This also includes any meetings where a plan may have been decided upon or abandoned or, indirectly, expressions of will that were related to appropriate plans were created by the ZK.

The user might also want to know, in a more general sense, whether the ZK had knowledge of the trip. In this case, the *Correspondence* pattern (V:2.1.4) could also be used to formulate queries targeting documents which refer to either a planning activity or related plan as described in this example, and which have been designated, addressed, or kept by the ZK or any member of this group. Documents which meet such a condition, especially documents which have been in the possession of the ZK, indicate a high likelihood that the ZK indeed had knowledge of the trip. However, once again, the historian must read and assess the documents in order to establish evidence and plausibility for this hypothesis.

### Example - “Aims of a group”

Q036-01-02<sub>barch</sub> “Which aims towards the GDR did the Greens pursue with their Deutschland-politik policy in the 1980s?”

The general question type is *resource discovery* <research question>. The wanted entity is

<group.activity>. The relevant *given* entities are *particular group* (“the Greens”, “GDR”), *particular activity* (“*Deutschlandpolitik* policy [of the Greens in the 1980s]”), and *particular period* (“the 1980s”).

This is a *psychological* question since it inquires as to intentions, an unobservable state of a group. The *provenance context* is *self-doc [particular group]* since the documents in question are the result of self-documentation activities conducted by the particular group “the Greens”. The *aboutness context* is *plans [particular activity]* since the question is generally about the context of a particular activity, “the [‘Deutschlandpolitik’] policy of the Greens towards Germany for the GDR in the 1980s”.

Primary information on the aims of the Greens can be expected especially in internal documents of the Greens, including particular plans and documents about plans or planning activities. Secondary information such as newspaper articles may also be relevant; however, while this sense is excluded in this exemplary query pattern, it could be easily represented using the *Documentation* pattern (V:2.1.2).

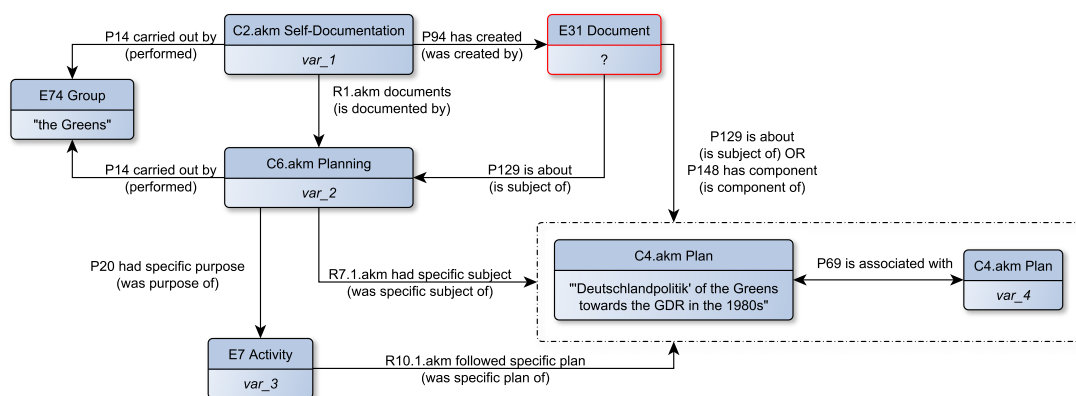


Figure 57 – Q036-01-02<sub>barch</sub>: Aims of a group.

The principle query pattern as shown in Figure 57 has the same structure as that shown in Figure 56. This shows once again how very different types of questions can be represented with one shared pattern. Relevant information objects would be:

1. any distinct plans related to the general “Deutschlandpolitik’ of the Greens towards the GDR in the 1980s” (C4<sub>akm</sub>)
2. documents (E31) either about such plans or which incorporate them, for example minutes of planning meetings
3. documents (E31) about any
  - (a) planning activities which either discussed the general “plan” “Deutschlandpolitik’ of the Greens towards the GDR in the 1980s” or any related plan,

- (b) or planning activities with a view to executing any plans in the context of the “Deutschlandpolitik” of the Greens.

These are same set of principal query patterns as in the example on page 190. The actual “aims” of the Greens are subject to the individual analysis of the historian studying the documents.

**Example - “Course of action of a group”**

Q036-01-03<sub>barch</sub> “What was the course of action of the Greens for achieving their objectives in the context of their Deutschlandpolitik towards the GDR?”

The general question type is *resource discovery* <research question>. The wanted entity is <group.activity>. The relevant given entities are *particular group* (“the Greens”, “GDR”), *particular activity* (“the Deutschlandpolitik [of the Greens in the 1980s]”), and *particular period* (“the 1980s”).

The question is a *collective statistical* question since it pertains to the general behaviour and conduct of a group. The *provenance context* is *self-doc* [particular group], since the most manifest interpretation the question refers to is the self-documentation of activities carried out by the Greens. The *aboutness context* is *plans* [particular activity] since the question is generally about the context of a particular activity, “the Deutschlandpolitik [of the Greens in the 1980s]”.

The question is very broad and demands several distinct query patterns in order to retrieve a sufficiently broad set of relevant documents. In contrast to the previous example, the focus is clearly on the “actions” of the Greens, less on the actual plans. The various query patterns would therefore focus on documents (E31) about activities (E7) carried out by the Greens (E74) following (R10.2<sub>akm</sub>) a particular plan “Deutschlandpolitik [of the Greens towards the GDR in the 1980s]” (C4<sub>akm</sub>) or any associated plan. This core pattern is shown in Figure 58.

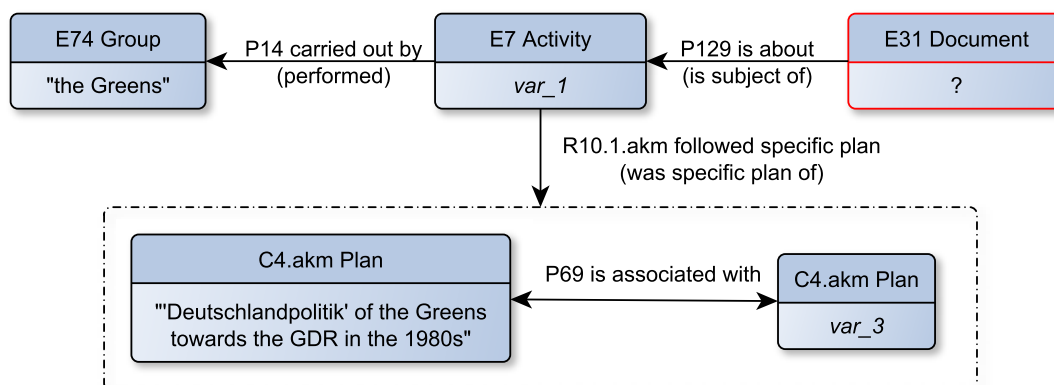


Figure 58 – Q036-01-03<sub>barch</sub>: Course of action of a group.

The different query patterns then mostly vary in terms of the provenance context of the documents about these activities or relevant plans.

Figure 59 covers cases where the Greens self-documented their activities conducted as part of their Deutschlandpolitik; for example, reports on meetings of members of the Greens with officials from the GDR.

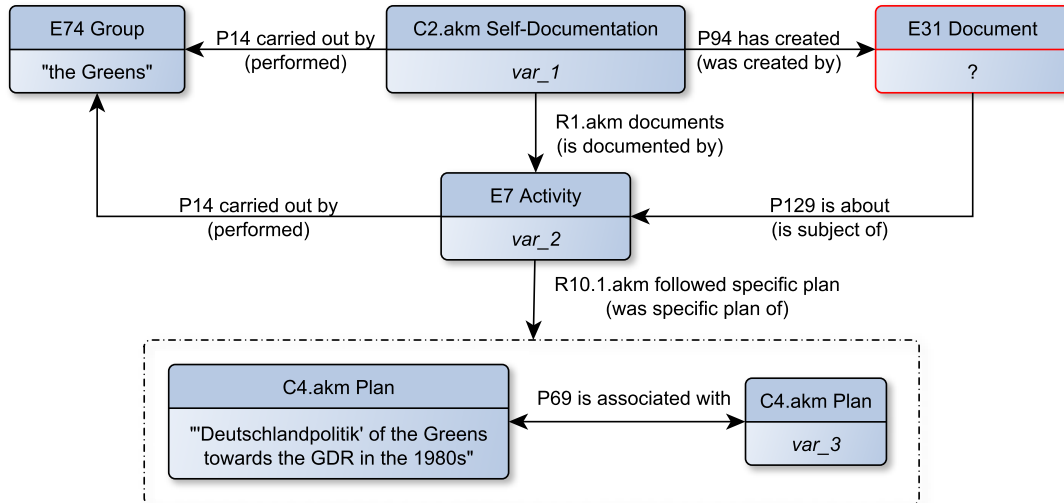


Figure 59 – Q036-01-03b<sub>barch</sub>: Course of action of a group.

The query pattern shown in Figure 57 in the previous example is a special case of the query pattern shown in Figure 59, covering instances related to any planning activities of the Greens in the context of their Deutschlandpolitik. This would include, as also previously discussed, the creation, decision, and abandonment of plans in the context of Deutschlandpolitik.

Figure 60 covers situations where the Greens indirectly document activities which occurred as a consequence of other activities conducted as part of their Deutschlandpolitik. Primary examples include the minutes (E31) of meetings following up on any activity. Sessions of the Bundestag may also talk about these activities of the Greens. In this case, the modelling of the actor who carried out the self-documentation, the session of the Bundestag, would be more detailed, including the “fraction of the Greens in the Bundestag” as a member of the group “Bundestag”. The minutes created by the self-documentation activity would not only be about the session of the Bundestag but also make statements about the activities of the Greens discussed during that session.

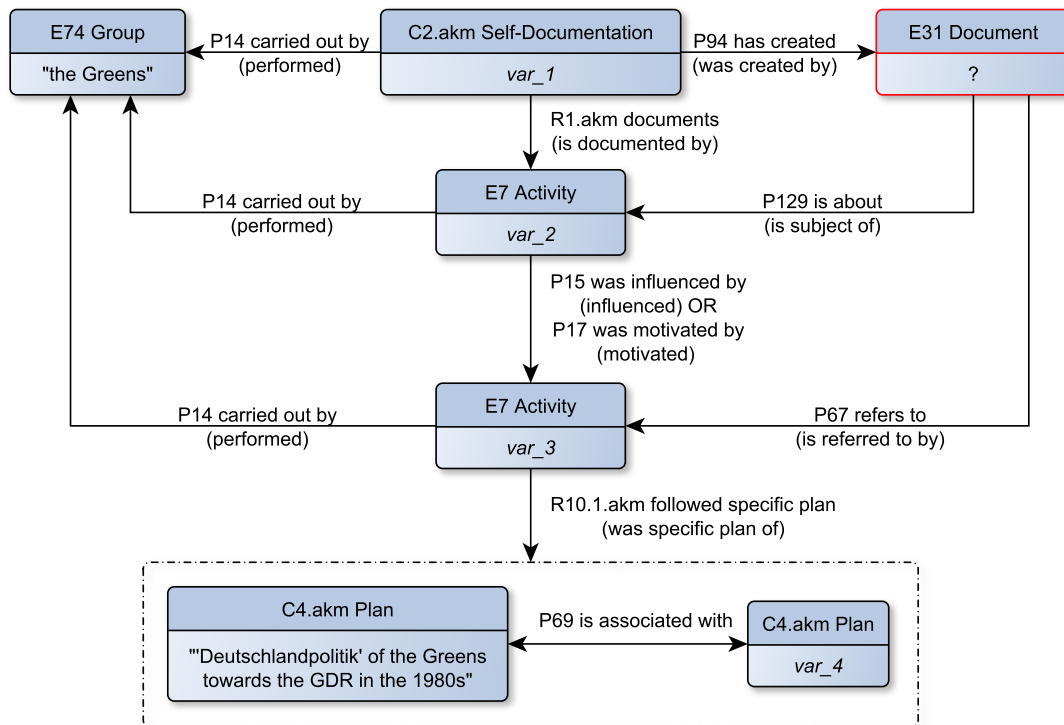


Figure 60 – Q036-01-03<sub>c</sub><sub>barch</sub>: Course of action of a group.

This query pattern is a mere variation of the previous one insofar as chains of activities are represented, as described in the *Events* pattern (V:2.2.2). Accordingly, cases of planning activities occurring as a consequence of other activities are conceivable.

The level of possible indirections depends on the richness of the knowledge base and, of course, on the interest of the user regarding inclusive or exclusive result sets.

Figure 61 shows the third possible query pattern describing situations where actors outside the particular group “the Greens” document (C1<sub>akm</sub>) the activities conducted by the Greens in pursuit of their “Deutschlandpolitik” (C4<sub>akm</sub>).

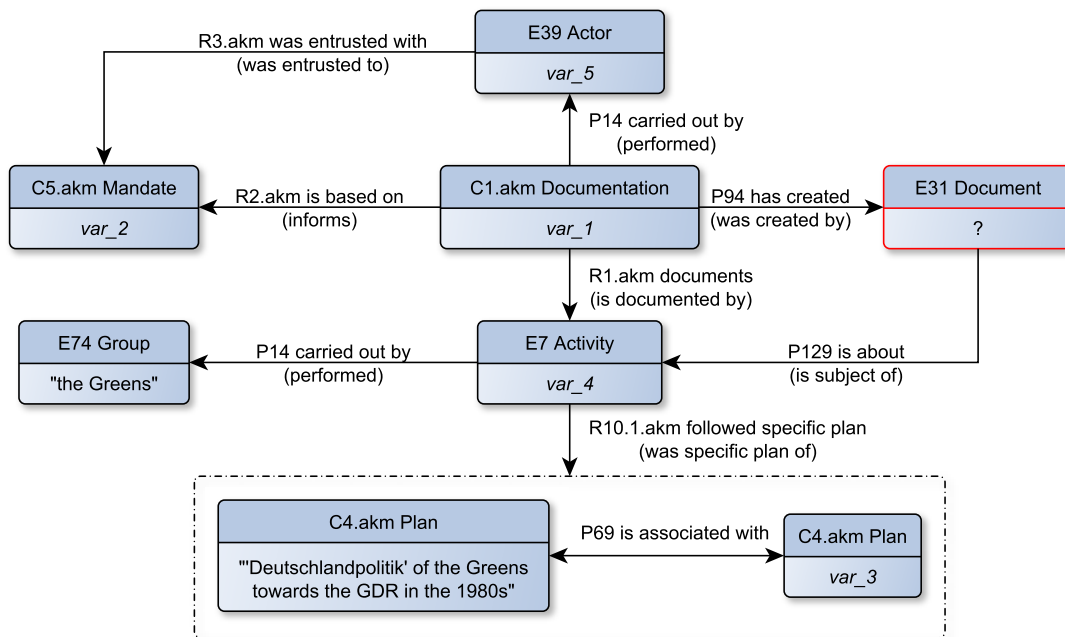


Figure 61 – Q036-01-03a<sub>barch</sub>: Course of action of a group.

For example, these could include any kind of news articles about the Greens and their activities in the context of Deutschlandpolitik.

Since the question is a collective statistical inquiry, any facts pertaining to the various query patterns are potentially equally relevant for the historian. For example, the historian might be interested in the temporal sequence of certain activities occurring in the context of the Deutschlandpolitik of the Greens.

Time and place, as well as types, have not been included in these exemplary query patterns.

### Example - “Decisions”

Q014-04-06<sub>barch</sub> “(...) no. 5.) 2nd March, 1933: [I am looking for the] *Decision of the Reichgovernment to host the opening of the Garrison Church in Potsdam. Yet without exact date.*”

The question at hand is item number five on a list with several questions for the location of specific items.

The general question type is *resource discovery* <specific item>. The *wanted* entity is <activity>. The relevant *given* entities are *particular document* (“decision [of the Reichgovernment to host the opening of the Garrison Church in Potsdam (2nd March, 1933)]”), *particular group* (“the Reichgovernment”), and *time-span* (“March, 2nd 1933”).



The question is a *material fact* question since it inquires as to an observable fact; that is, the existence of a particular document. The *provenance context* is *self-doc [particular group]* since the question asks about a decision taken by the particular group Reichgovernment as the result of a self-documentation activity. The *aboutness context* is *plans [particular activity]* since the information object asked for, the decision, refers to a particular activity.

While the example on page 192 mentioned expression of will only as a possible indirect query target, the query pattern shown in Figure 62 explicitly represents a particular instance of expression of will.

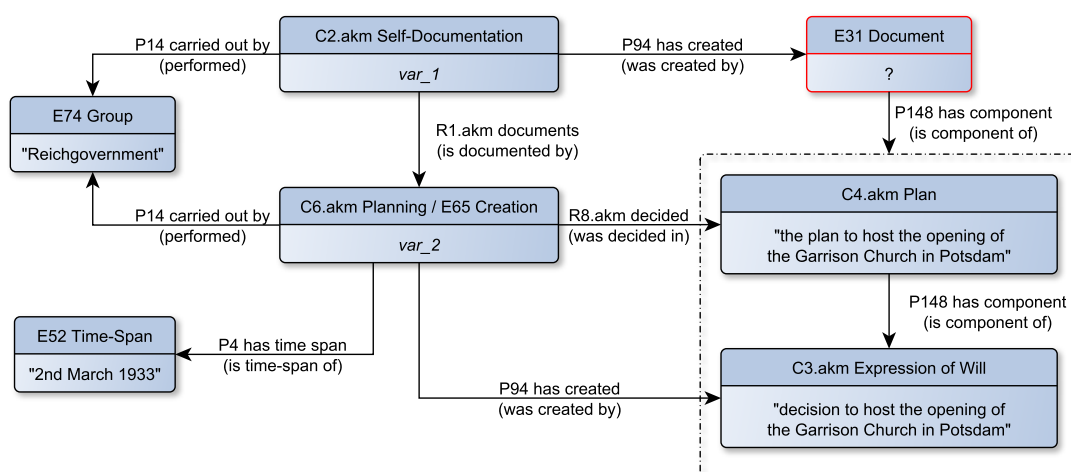


Figure 62 – Q014-04-06<sub>barch</sub>: Decisions.

The particular group Reichgovernment (E74) carried out an activity (E65) which “created” the “decision to host the opening of the Garrison Church in Potsdam” (C3<sub>akm</sub>), for example during a meeting of the cabinet. This meeting was documented (C2<sub>akm</sub>) by the Reichgovernment (E74) itself by creating a document (E31), such as minutes, which would include (P148) the decision in question. The decision was taken on March 2nd 1933.<sup>84</sup>

In theory, albeit unlikely, the Reichgovernment might have conducted a meeting where only this particular decision was taken and recorded, in which case the decision would not be a component of another document but would only be this document. In other words, the decision would be an instance of C3<sub>akm</sub> *Expression of Will* and of E31 *Document* resulting from a planning activity (C6<sub>akm</sub>) and a self-documentation activity (C2<sub>akm</sub>).

If an instance of the class C3<sub>akm</sub> *Expression of Will*, here the “decision to host the opening of the Garrison Church in Potsdam”, was created during an activity (E65), then this activity must also have been a planning activity (C6<sub>akm</sub>) since its participants must have talked (R7.1<sub>akm</sub>)

<sup>84</sup> The date mentioned in the inquiry, 2nd March 1933, most likely refers to the date on which the decision was taken. Note here too that the self-documentation activity may also take place after the documented activity has finished.

about a particular plan ( $C4_{akm}$ ). In the current example, this planning activity ( $C6_{akm}$ ) must have decided ( $R8_{akm}$ ) to execute the “plan to host the opening of the Garrison Church in Potsdam” ( $C4_{akm}$ ). The particular plan may therefore include the decision or both information objects may be included in the minutes.

### Example - “Having the mandate to do something”

Context “Sources I would like to consult are police and surveillance reports from the Weimar Republic about revolutionary movements. (...)”

Q002-03-03<sub>barch</sub> “Which agency of the Reich was responsible for the surveillance of the revolutionary movements?”

The general question type is *fact-finding*. The *wanted* entity is  $\langle group.identification \rangle$ , which subsumes interests in mandates. The relevant *given* entities are *type of activity* (“surveillance”), *particular group* (“the Reich”), *type of group* (“revolutionary movements”), and *particular period* (“Weimar Republic”).

The question is a *material fact* question. The *provenance context* is *documentation [actor]* since the particular actor who had the mandate to conduct surveillance is the actual primary target of the question. The *aboutness context* is *plans [type of activity]* since the historical background of the question refers to “surveillance” as a type of activity.

The primary wanted information is the name of the actor entrusted with the mandate to carry out documentation activities of the type “surveillance” which document or observe the activities of a type of actor, and “revolutionary movements” during the period of the Weimar Republic.

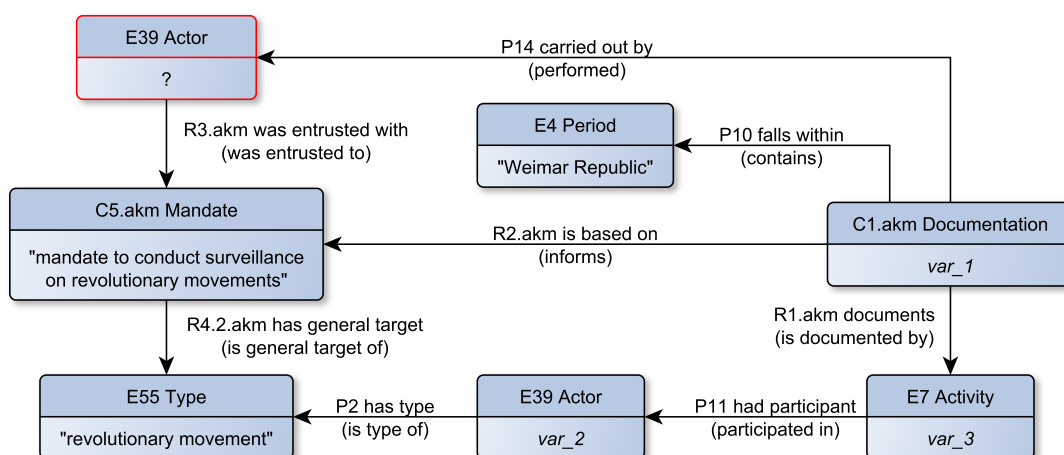


Figure 63 – Q002-03-03<sub>barch</sub>: Factual questions about the name of an actor entrusted with a specific mandate.

Figure 63 shows an exemplary instantiation of a query pattern. The word “responsible” is important because it stresses the fact that whichever agency conducted the surveillance activities did so following a mandate which formally delegated said responsibility to the agency.

In cases of factual questions, a query would first seek to retrieve the name from a given knowledge base. However, if the fact is not discovered and since archival knowledge bases are not primarily intended for factual retrieval and historians would probably demand written evidence for the retrieved fact, a second query is always assumed which targets documents adequate to provide potential phenomenological evidence.

The actor participating in a documented activity may be a group or a person as a member of a group. Figure 63 therefore shows a second query target, the document incorporating the mandate. Any factual question could thus be considered a potential and implicit resource discovery question.

**Statistics** In the complete sample, the *Plans* pattern has been assigned to 70 questions representing 15% of all inquiries, of which 63 belong to the BArch sample representing 18% of all inquiries in that sample and 7 to the NAN sample representing 5.3% in that sample. Within the context of the *Plans* pattern, only in five cases (7.1%) could the particular focus not be reasonably determined, of which only one (1.6%) occurred in the BArch sample and the remaining four (57%) in the NAN sample. The most frequent primary entities of interest in the *Plans* pattern are type of activity with 44 counts (63%), of which most pertain to the BArch sample with 41 counts (65%), and only 3 (43%) to the NAN sample. In 21 cases (30%) particular activities were the primary entity of interest, all of which are part of the BArch sample. Clearly, in the context of plans and planning, most users were interested in information objects about types of activities. In the case of the *Events* pattern (V:2.2.2) this ratio is reversed.

<i>Aboutness Context</i>	<i>All</i> (n=467)		<i>BArch</i> (n=345)		<i>NAN</i> (n=131)	
<b>Plans</b>	70	15%	63	18%	7	5.3%
- Plans [ any ]	5	7.1%	1	1.6%	4	57%
- Plans [ particular activity ]	21	30%	21	33%	0	-
- Plans [ type of activity ]	44	63%	41	65%	3	43%

Table 12 – Occurrences of the general pattern *Plans*. Note that percentages for the primary entities of interest are per current general pattern.

**Summary** The *Plans* pattern extends the *Events* pattern (V:2.2.2) by the two basic notions of planning and wanting of activities by actors. The pattern specifically allows us to describe the planning of activities and their plan-based execution, more specifically the discussion and creation of plans, and the principle beginning and end of a will to execute a plan.

The pattern particularly distinguishes between the information objects plan, expression of will, and document. This distinction allows for questions about plans and enactments of plans to be generally addressed, and in particular questions related to the intentions and aims of actors, to what extent plans were applied and executed, and who desired the specific plans; that is, intentions to execute plans versus the reality of the actual execution and enforcement.

With regard to archival records and archival description, the *Plans* pattern further explicitly introduces a temporal perspective looking into the future from any given point in the past referred to by the record or the archival material. The pattern represents and describes what was planned, intended, and expressed as will in order to compare these plans and intentions with what is known or said, has been recorded.

The CRM exhibited the most significant shortcomings for the purposes of this study in the context of the *Plans* pattern. A couple of new classes and properties therefore had to be introduced in order to cover the aforementioned semantics: The class *C4<sub>akm</sub> Plan* specifically represents plan structures as either a description of appearance such as topographical maps, formal procedures such as laws, or intended, projected future states such as agendas or social policies. The class *C6<sub>akm</sub> Planning* then covers activities preparing and discussing such plans and the psychological entity *C3<sub>akm</sub> Expression of Will* represents the will to enact, execute, or realize them.

With these three classes, several material facts related to the described framework of plan and will can be created, utilizing, amongst other aspects, a range of new properties: *R10.1<sub>akm</sub> followed general plan (was general plan of)* and *R10.2<sub>akm</sub> followed specific plan (was specific plan of)*. These express whether an activity, specific or general, was based on or followed a particular plan.

The act of planning a plan may at the same time be an act of creation of that plan and thus entail an expression of will which may or may not be directly or indirectly recorded. Similarly, planning activities may decide upon (*R8<sub>akm</sub> decided (was decided in)*), or abandon (*R9<sub>akm</sub> abandoned (was abandoned by)*) an existing plan and at the same time create an expression of will. For example, the minutes of a planning meeting may contain the formal remark of having accepted a plan. Planning activities may further discuss plans, either types of plans (*R7.2<sub>akm</sub> had general subject (was general subject of)*) or particular plans (*R7.1<sub>akm</sub> had specific subject (was specific subject of)*).

#### 2.2.4 Actors

Ultimately, all history is about and carried out by individuals and groups. Actors have been represented in all general patterns so far; the *Actors* pattern consequently details the various essential constituents relevant to the being, acting and suffering of historical actors.

**General Pattern** The pattern *Actors* represents factual questions about actors in general and things about actors and their activities. Questions belonging to this pattern focus on obtaining

either facts or documents about one or more actors and their lives. The following principal senses can be distinguished:

1. Identifiable but *immaterial* things (*E73 Information Object*) about a person (E21) or a group (E74).
2. Identifiable but *immaterial* things (*E73 Information Object*) that are about the activities carried out by a person (E21) or a group (E74).
3. Any *facts* about a person (E21) or group (E74) such as the date of birth of a person, the membership of a person, or the existence of a group.

Figure 64 shows the general *Actors* pattern. These questions are often factual and have a strong *biographical* notion. Examples include general questions asking about any biographical information about a particular person, or specific questions about the membership of a person in a group.

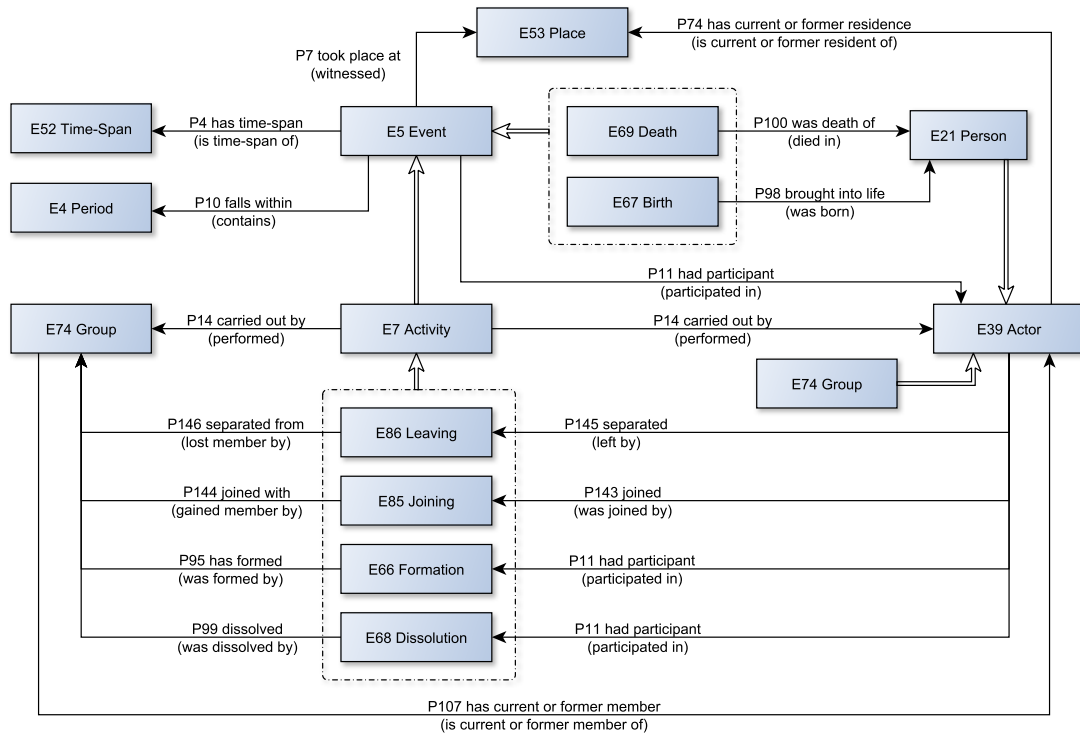


Figure 64 – General *Actors* pattern.

In contrast to the *Events* pattern (V:2.2.2), the primary focus of the *Actors* pattern is on the actor that is, a person (E21) or group (E74), and the activities (E7) intentionally carried out (P14) by this actor. The *Events* pattern rather stresses the passive role of an actor and is primarily interested in the event or activity.

The *Actors* pattern allows for the representation of any activity (E7) during which an actor has been either actively or passively present (P11) or which has been carried out by (P14) the actor. For example, a soldier participates (P11) in a war while the writing of a book is carried out (P14) by the author.

Many questions, such as general biographical questions, inquire as to documents about a particular actor, mostly persons (E21), and any event or activity during which this actor was either present (P11) or actively involved (P14). An instructive example of this typical application is given at the end of this section.

Another important principal application of the *Actors* pattern pertains to questions about the relationship between a person and a group. Such questions evolve around two primary sets of particular types of activities:

1. the *joining* (E86) or *leaving* (E85) of groups (E74)
2. the *formation* (E66) or *dissolution* (E68) of groups (E74)

The concept of groups in the CRM, as discussed previously (IV:3.2), comprises groups of people as any gatherings or organizations of more than two people, including the concept of the family and the office. The membership of persons (E21) in groups (E74), especially political parties and families, is an important aspect in the questions either as the wanted information or as given contextual information.

For example, many questions ask if and when a person was a member of a particular political party. The fact that a person (E21) is a member in a political party (E74) can be represented by the property *P107 has current or former member (is current or former member of)*.

The joining and leaving of a group can be represented by the classes *E85 Joining* and *E86 Leaving*. For example, the joining of a political party can be represented by an instance of the class *E85 Joining* which joins (*P143 joined (was joined by)*) a person (E21) with (*P144 joined with (gained member by)*) the political party (E74). Accordingly, termination of the membership can be represented by an instance of the class *E86 Leaving* which separates (*P145 separated (left by)*) the person (E21) from (*P146 separated from (lost member by)*) the political party (E74).

Marriages can be modelled in the same way, where two persons (E21) form a new group: the family. Children would become members (P107) of that group (E74) via an instance of *E85 Joining* which would be at the same time an instance of the class *E67 Birth*. Accordingly, natural death would be represented by an instance of *E86 Leaving* which, at the same time, would be an instance of the class *E69 Death*.

Constitution and dissolution of groups such as political organizations, nations, but also families, can be modelled using the classes *E66 Formation* and *E68 Dissolution*. While persons or groups participated (*P11 had participant (participated in)*) in these activities, instances of *E66 Formation* form (*P95 has formed (was formed by)*) and instances of *E68 Dissolution* dissolve (*P99 dissolved (was dissolved by)*) instances of groups (E74).

Since all these classes are sub-classes of *E7 Activity*, they can be further described by the place (E53) where they happened and the date (E52) or period (E4) on or during which they occurred.

Unintentional events (E5), as has been discussed in the context of the *Events* pattern (V:2.2.2), are not relevant in the context of this analysis, except for cases of birth and natural death events.

The places and dates of the birth and natural death of a person (E21) constitute important *given* information and imply events of the type *death and birth*. Such events are represented by the classes *E67 Birth* and *E69 Death*. They are further described by the place (E53) where the birth or death took place (*P7 took place at (witnessed)*) and the date (E52) (*P4 has time-span (is time-span of)*). The person (E21) who was born or who died during these events is connected by the properties *P98 brought into life (was born)* and *P100 was death of (died in)*.

The place of residence of a group or person was given in many questions from the sample. The fact that an actor (E39) has been resident at a particular place (E53) can be expressed by the property *P74 has current or former residence (is current or former resident of)*. This kind of information helps to identify groups (E74) which may have created documents (E31) about a person (E21), for example local authorities or dependencies of government agencies.

As the following examples will demonstrate, it is not often easy to confidently conclude from the utterance and the context of a question whether the user is primarily requesting documents about an actor, in most cases particular persons, or if the user also requires documents created or produced by the actor. Once again, the categorization of the question via general patterns from the provenance and aboutness contexts, reflects a perceived primary epistemic focus of the question rather than a definite selection.

### Example - “Documents about a type of group”

Q059-05-01<sub>barch</sub> “Does your archive have holdings of this organizational unit of the NSDAP (Reich leadership. Delegate of the Führer for the entire ideological instruction of the NSDAP) with documents [about] philatelic organizations?”

The general question type is *resource discovery <material-finding>*. The *wanted* entity is *<group.regarding>*. The relevant *given* entities are *particular group* (“Reich leadership. Delegate of the Führer for the entire ideological instruction of the NSDAP”) and *type of group* (“philatelic organizations”).

This is a *material fact* question since it inquires about an observable subject relationship between documents and a type of group. The *provenance context* is *provenance [particular group]* since any document kept or created by the aforementioned group is potentially relevant. The *aboutness context* is *actors [type of group]*.

Figure 65 shows the obvious and most generic query pattern for this question. The query asks for any kind of textual or visual information object (E73) referring to the type of group “philatelic organizations” and which has been either created or kept by the particular group “Reich leadership. Delegate of the Führer for the entire ideological instruction of the NSDAP”.

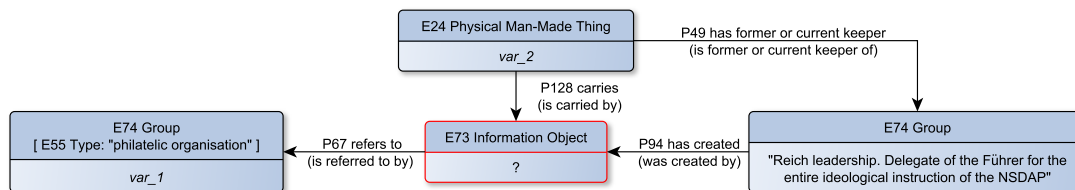


Figure 65 – Q059-05-01<sub>barch</sub>: Documents about a type of group.

This question is a simple and basic example of a question pertaining to the *Actors* pattern. The query instantiation could be reasonably adjusted, as has been discussed in many previous cases. For example, the query could specifically ask for documents (E31) which have philatelic organizations as a topic (P129), or for reports (E31) from documentation activities (C1<sub>akm</sub>) about (P129) the various activities of philatelic organizations such as member gatherings or exhibitions organized by such groups.

#### Example - “Any papers regarding a particular person”

Context            “I am interested in archive material involving the former leader of the Danish Communist Party, [person name]. He was leading the party from [date]-[date].”

Q010-01-01<sub>barch</sub> “Can you investigate whether there are more papers in your archive regarding [person name]?”

The general question type is *resource discovery <material-finding>*. The wanted entity is *<person.regarding>*. The relevant given entities are *particular person* (“[person name]”). The *supplementary* information includes *particular group* (“Danish Communist Party”), *type of group* (“leader of the Danish Communist Party”), *type of activity* (“leadership”), and *time-span* (“[date]-[date]”).

This is a *material fact* question since it inquires about an observable subject relationship between a document and a person. The *provenance context* is *provenance [actor]*. The *aboutness context* is *actors [particular person]*.

The most likely interpretation of the question is that the user is seeking anything at all pertaining to the particular person, “[person name]”, and not documents created or kept by this person. However, the latter interpretation would be representable using the *Provenance* pattern (V:2.1.1). In this case, the aboutness context would be *aboutness [any]* and the provenance context would be *provenance [particular person]*.

Figure 66 shows the most basic query pattern for the question asking for any visual or textual information object (E73) containing a statement (P67) about “[person name]”. Since the provenance is open, the notion of *possession* of things related to “[person name]” is not required.



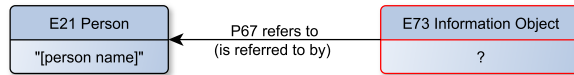


Figure 66 – Q010-01-01<sub>barch</sub>: Any things about a particular person.

Figure 67 shows an extended query pattern with all the contextual information given in the question. This extended example demonstrates how membership and offices are modelled in the CRM.

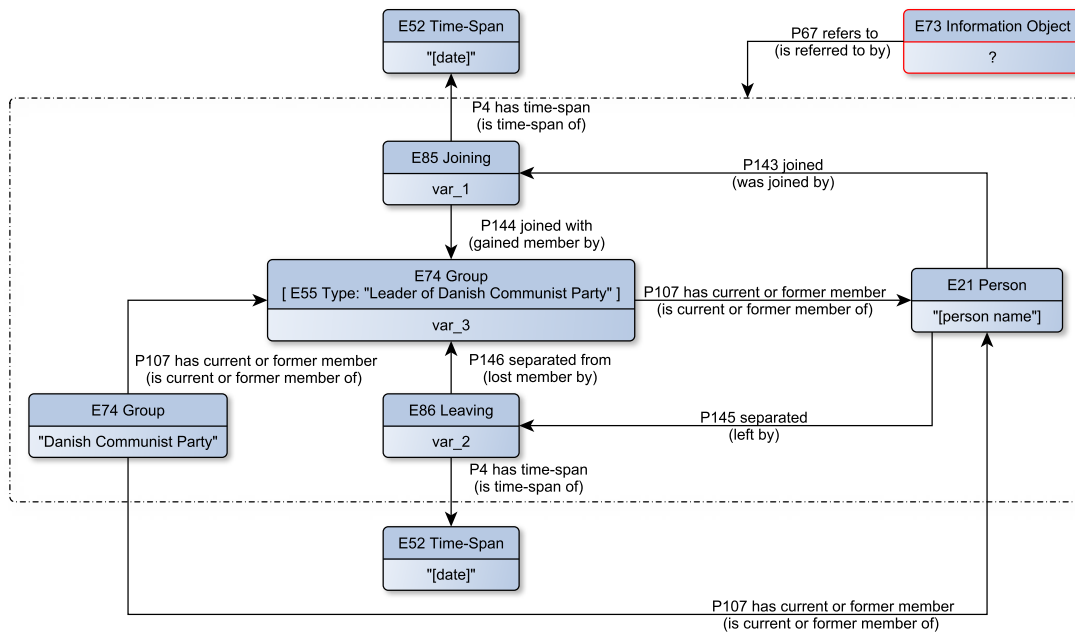


Figure 67 – Q010-01-01<sub>barch</sub>: Any things about the activities of a particular person.

Especially in cases of very general or unspecific questions, many other implicit variations of the patterns are possible. For instance, a query might inquire into documents about any activity carried out by “[person name]”, possibly as a result of a documentation activity.

### Example - “Membership of a particular person”

Q003-05-01<sub>barch</sub> *“I would like to know whether and when the spouses of individual communist intellectuals joined the KPD, [person name] (born [maiden family name], date of birth and place of birth unknown).”*

The general question type is *fact-finding*. The *wanted* entity is *<activity.date>*. The relevant *given* entities are *particular person* (“[person name]”), and *particular group* (“KPD”).

The question is a *material fact* question since it inquires about an observable relationship between a person and a group. The *provenance context* is *documentation [particular group]*. The *aboutness context* is *actors [particular person]*.

Figure 68 provides the general query pattern with the relevant context for several possible queries. The question as to *whether* the particular person “[person name]” joined the political party KPD or not could be answered, for example, by an existential query for a relationship *P143 joined (was joined by)* between the *E21 Person* “[person name]” and the particular activity *E85 Joining* “Joining KPD” or for a relationship *P107 has current or former member (is current or former member of)* between the particular actors *E74 Group* “KPD” and *E21 Person* “[person name]”.

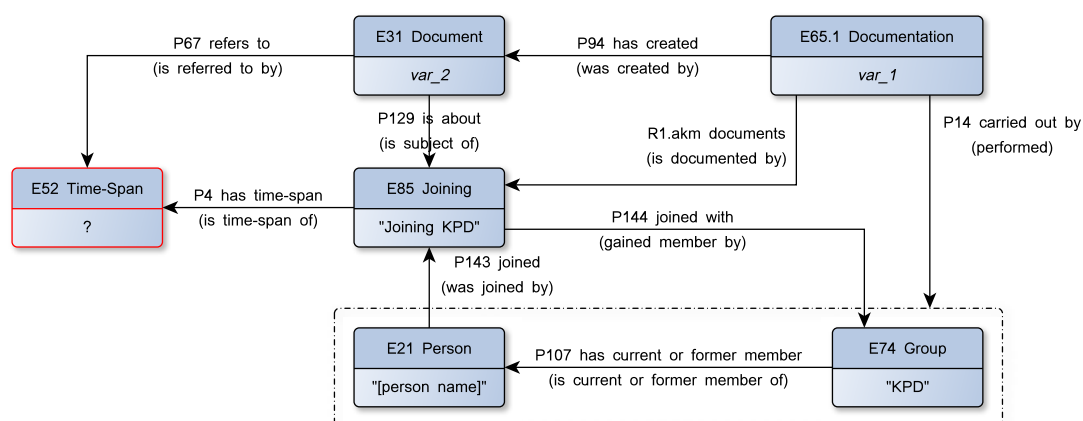


Figure 68 – Q003-05-01<sub>abarch</sub>: Membership of a person in a political party.

In order to obtain evidence in the form of a document, the query pattern could target a documented joining activity, either by the KPD or by the person itself. In the latter case, this would be a self-documentation. What is not shown in this figure is that of course any other actor could have documented this joining activity as well; however, but the principal query pattern remains the same.

The actual date of joining could either be retrieved as a fact directly from the semantic network, if recorded, or otherwise indirectly from one of the documents retrieved which mentions the joining activity.

Another possible and particular query pattern targets the party membership card as a tangible and specific document. Figure 69 represents the context of a query pattern in search of the party membership card in the personal estate of “[person name]”.

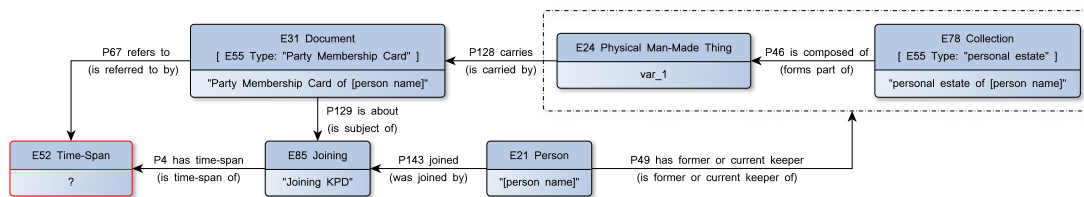


Figure 69 – Q003-05-01b<sub>barch</sub>: Party membership card of a person.

Another indirect query pattern would be for any documented leaving activities which, if found, would lead to the conclusion that “[person name]” must have been a member of the KPD.

This example demonstrates once again that there are often several possible queries suited to satisfy the information need. Even though the model is highly flexible and efficient, query patterns often boil down to the same basic patterns of relationships and entities.

**Statistics** In the complete sample, the *Actors* pattern has been assigned to 207 questions representing 43% of all inquiries. Of these 207 questions 153 belong to the BArch sample representing 44% of all inquiries in that sample, while the remaining 54 questions belongs to the NAN sample representing 41% of inquiries in that sample. Within the *Actors* pattern, the by far most dominant primary entity of interest is *particular person* with 161 occurrences (78%) of which 124 (81%) are from to the BArch sample and 37 (69%) from the NAN sample. *Particular groups* come second, amounting to 23 counts (11%) in the context of the *Actors* pattern, of which 16 (10%) are part of the BArch sample and 7 (13%) of the NAN sample. Types of groups and types of persons are far less relevant with 17 counts (8.2%) for the former and 6 counts (2.9%) for the latter.

Aboutness Context	All (n=467)		BArch (n=345)		NAN (n=131)	
<b>Actors</b>	207	43%	153	44%	54	41%
- Actors [ particular person ]	161	78%	124	81%	37	69%
- Actors [ type of person ]	6	2.9%	2	1.3%	4	7.4%
- Actors [ particular group ]	23	11%	16	10%	7	13%
- Actors [ type of group ]	17	8.2%	11	7.2%	6	11%

Table 13 – Occurrences of the general pattern *Actors*. Note that percentages for the primary entities of interest are per current general pattern.

**Summary** The *Actors* pattern covers persons and groups, their lives and histories. The pattern does include activities; however, in contrast to the *Events* pattern (V:2.2.2), the focus is on the history of the actor and the activities the actor carried out. The *Actors* pattern can thus be considered biographical, covering the typical activities in the lives of people such as joining

various kinds of groups, including marriage, birth and death, and the various relationships and encounters between people and groups.

The *Actors* pattern covers open biographical questions about any aspect of the lives of persons or factual questions about specific aspects of people’s lives. Particularly important were questions about the membership of persons in various kinds of groups, especially political parties.

## 2.2.5 Things

In history, not only persons and groups in the contexts of activities and events are important but also the physical man-made things produced and utilized.

**General Pattern** The general pattern *Things* covers questions related to physical things fabricated by human activity, and allows for coarse descriptions of important events in their general history.

The pattern is based on the interpretative analysis of a relatively small set of broad and mostly material-finding questions inquiring about the general history of things. The wanted entities were, in all cases, particular things such as particular submarines, ships or buildings. Even though the general pattern *Things* has relatively small statistical relevance in the sample with overall only 11 occurrences in the NAN sample, inquiries about facts and documents related to various historical objects appears to be generally relevant, including for actor-centric historical inquiry.

Figure 70 shows the general pattern *Things*. Most of the constituents of this general pattern have already been introduced and discussed in the context of other general patterns.

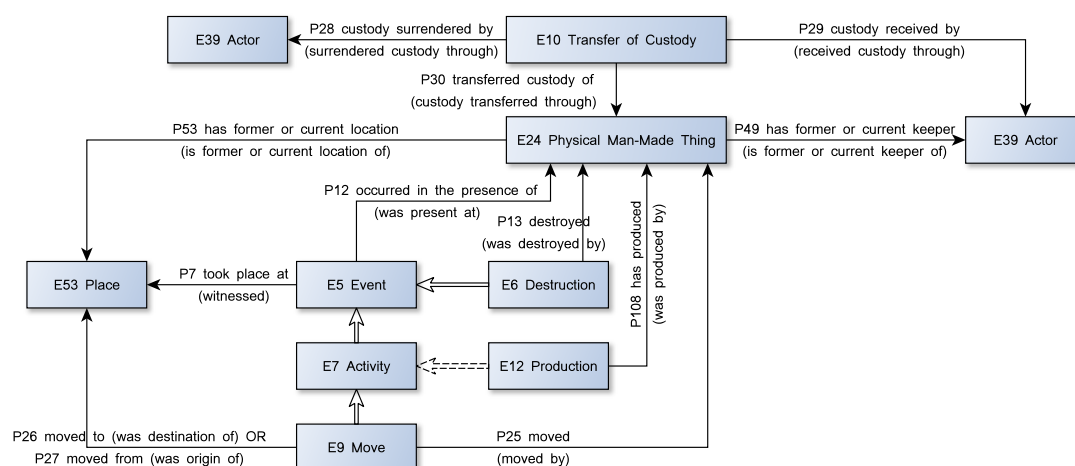


Figure 70 – General pattern *Things*.

The class *E24 Physical Man-Made Thing* is at the centre of this general pattern. As already discussed in the context of the general pattern *Provenance* (V:2.1.1), the class represents any kind

of physical thing which has been deliberately produced by human activity. In the context of the current pattern, these things may include submarines, buildings or ships.

The history or fate of these things can be described by placing them into the context of events (E5) or activities (E7) using the property *P12 occurred in the presence of (was present at)*. Most of the further essential properties for events and activities have already been introduced in the context of the general pattern *Events (V:2.2.2)* and have therefore been omitted from Figure 70 for the sake of readability. Both activities and events may take place at (P7) geographical locations (E53), fall within (P10) historical periods (E4), and occur (P4) during time-spans (E52). The essential pattern of involvement of actors (E39) in events and activities has also been omitted. However, persons or groups may either actively or passively participate (P11) in events (E5) or carry out (P14) activities (E7).

When investigating the general history of mobile things of “utility”, in the sample used for this study important questions inquired after particular submarines and ships, and the places where these vessels were in dock during their lifetimes, including the current or last known whereabouts. The current and past whereabouts of things is represented not only indirectly through events and activities taking place at particular places but also directly by the property *P53 has former or current location (is former or current location of)*. By using this property, places which have been home to specific things can be easily documented without regard to the involved events, time-spans, or actors. The places where a thing has been or has been used or employed may already provide valuable information regarding other archives where further information and documents may be found.

The property *P53 has former or current location (is former or current location of)* is actually a shortcut for the longer and more informative path using the class *E9 Move*. The class allows for detailed descriptions as to which things have been moved (P25) from one place to another using the properties *P25 moved (moved by)*, *P26 moved to (was destination of)*, and *P27 moved from (was origin of)*. Since *E9 Move* is a sub-class of *E7 Activity*, the other properties introduced in the context of the general pattern *Events (V:2.2.2)* may also be used. For example, the actors involved in the moving may be specified, or the time-span of the moving process. Furthermore, if necessary, the activity of moving itself may be further described as a sequence of activities or as consisting of several sub-activities.

Two additional sub-classes of *E7 Activity* appear to be relevant: *E12 Production* and *E6 Destruction*, denoting the beginning and ending of the existence of the things to which inquiries refer. The class *E12 Production* has been discussed in the context of the general pattern *Provenance (V:2.1.1)* and covers the activity of producing a thing such as the building of a submarine or ship. The class *E6 Destruction*, on the other hand, describes the end of the existence of a thing in its intended form, role or character. For example, a submarine or ship which is destroyed by an attack at sea ceases to be functional and able to fulfil its intended purpose, normally accompanied by a more-or-less extensive modification of its physical form such as holes in the ship’s hull. In this regard, the question as to when an activity can be considered as an instance

of *E6 Destruction*, expressed by the property *P13 destroyed (was destroyed by)*, is rather one of interpretation and of the intended purpose of the documented information about the thing itself.

The class *E10 Transfer of Custody* and its properties will be discussed in more detail in the general pattern *Documents (V:2.2.6)*. As a sub-class of *E7 Activity*, *E10 Transfer of Custody* is a special kind of activity and permits the representation of a broad range of activities evolving around the change of ownership or custody over things in the widest sense, for example the theft of an object. At least two actors (*E39*) may be involved – neither of whom needs to be known – where one actor surrenders custody voluntarily or involuntarily, expressed by the property *P28 custody surrendered by (surrendered custody through)*, and another actor receives custody, expressed by the property *P29 custody received by (received custody through)*. The thing involved in the transfer of custody is connected to the activity by the property *P30 transferred custody of (custody transferred through)*. A particular example of transfer of custody outside of the archival context, which will be discussed in the context of the general pattern *Documents (V:2.2.6)*, is the acquisition of German submarines by Norway after the end of the Second World War. If the activity of transfer of custody also involves the moving of the thing to a new location, then this activity is also an instance of the class *E9 Moving*.

In the following, two examples will demonstrate the role of the general pattern *Things* in representing adequate query patterns.

#### **Example - “The fate of submarines”**

Context	<i>“I already have (...) a copy of a RNoN document dated 14 May 1948 authorizing funds for the repair and overhaul of the three Type VIIC submarines at Trondheim, but ...”</i>
Q013-02-03 <sub>nan</sub>	<i>“... I have very little information concerning their [the submarines’] fate during the intervening period.”</i>

The *question type* is *resource discovery <material-finding>* since the user is most likely seeking any kind of information objects providing details of what happened to the submarines. The *wanted entity* is *<thing>* since the interest of the inquiry focuses on the three submarines. The relevant *given entities* include *time-span* (“1945-1948”), *type of thing* (“submarine type VIIC”), and also *particular place* (“Trondheim”) as potentially relevant geographical information.

This is a *material fact* question since it inquires about an observable relationship between an information object and things. The *provenance context* is *documentation [actor]* since the relevant information objects must have been created during a documentation activity carried out by unknown actors. The *aboutness context* is *things [particular thing]* since the question and the demanded information objects refer to three particular submarines of the type VIIC, and not to any submarines of this type.

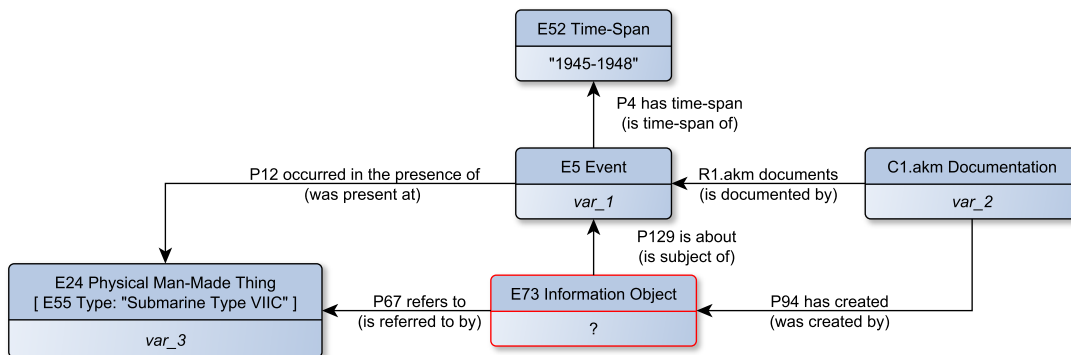


Figure 71 – Q013-02-03<sub>nan</sub>: The fate of submarines.

Figure 71 shows a query pattern covering a range of possible queries adequate to serve the interest of the current inquiry. As usual with very broad and unspecific questions, there are several potential formulations for such queries. Ultimately, queries will inquire as to any kind of information object (E73) with the topic of or at least reference to submarines (E24), about which type information may also be known (E55), or about any events or activities involving such submarines. Such queries could be further limited to particular sub-classes of activity (E7), such as the transfer of custody, moving or destruction. The time-span “1945-1948” could also further limit the activities involving these vessels, a time-span derived from the expression “intervening period”.

As in the case of many other previous examples, additional background knowledge and educated assumptions will help to further specify adequate queries such as by identifying possible authorities (E39) and mandates (C5<sub>akm</sub>). Potential results of such queries will provide a basis for further investigation by the inquirer regarding the fate of the submarines.

#### Example - “Information about two villas”

Context                    *“The Ministry has stated that it plans to publish a book about the Embassy Residences (Inkognitogata 5) and Embassy Chancery (Inkognitogata 7).”*

Q047-01-01<sub>nan</sub>        *“In this context, the Embassy has been asked to provide the as much information as possible (history, pictures, etc ...) on the two villas.”*

The *question type* is *resource discovery* <material-finding> since the user is seeking any kind of information related to the two villas. The *wanted entity* is <thing> since the interest of the inquiry focuses on two particular buildings. The relevant *given entities* include a kind of *identifier* (“Inkognitogata 5”, “Inkognitogata 7”), and *particular thing* (“villa”).

This is a *material fact* question since it inquires about an observable relationship between an information object and things. The *provenance context* is *provenance [actor]* since the context of creation of potentially relevant sources of information is unknown and the inquiry should

therefore remain inclusive. The *aboutness context* is *things [particular thing]*, since the question and the information objects requested refer to three particular buildings, the Embassy Residences (Inkognitogata 5) and Embassy Chancery (Inkognitogata 7).

Figure 72 shows a basic query pattern which could be the starting point for identifying relevant information objects and factual information on the two villas.

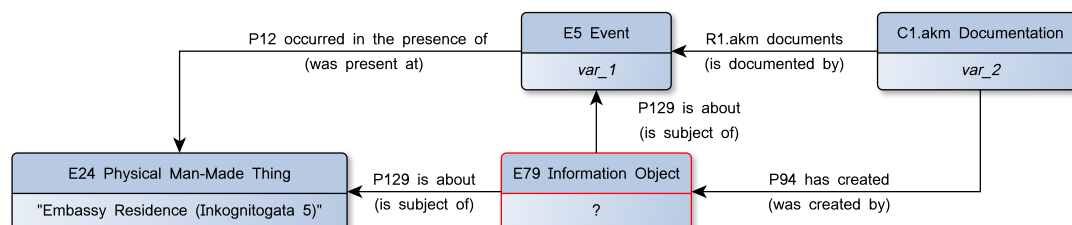


Figure 72 – Q047-01-01<sub>nan</sub>: Information about two villas.

As in the previous example, the question remains broad and inclusive, and information derived from background knowledge such as the place where the villas are located will support the specification of potential queries.

**Statistics** The *Things* pattern only appears in the inquiries collected from NAN in 15 cases representing 11% of all question in the sample, all of which refer to *particular things* as the primary entity of interest. In the context of the complete sample, the *Things* patterns only represents 3.2% of all questions.

Aboutness Context	All (n=467)		BArch (n=345)		NAN (n=131)	
<b>Things</b>	<b>15</b>	<b>3.2%</b>	0	-	15	11%
- Things [ particular thing ]	15	100%	0	-	15	100%
- Actors [ type of thing ]	0	-	0	-	0	-

Table 14 – Occurrences of the general pattern *Things*. Note that percentages for the primary entities of interest are per current general pattern.

**Summary** The *Things* pattern covers basic but relevant constituents of a general history of physical man-made things. While the empirical evidence is not strong in the current sample, the pattern can, however, be considered important for historical inquiries generally addressing the history of objects or as secondary questions in the contexts of other inquiries. The general pattern *Documents*, which will be discussed in the next section, can be seen as an extending sub-pattern of the general pattern *Things* focusing on particular kinds of physical man-made things; that is, traces of human activity and their properties, usually in written form.



## 2.2.6 Documents

While all patterns so far have primarily focused on information objects outside the archive, the *Documents* pattern also describes the textual and visual contents of physical man-made things within the context of an archive.

**General Pattern** The pattern *Documents* covers questions about the *context of physical and immaterial things* and their various qualities and attributes in the context of the archive. In contrast to the *Things* pattern (V:2.2.5), in this case, “things” refers to archival materials, primarily textual documents, but also pictorial, audio and audiovisual documents and their carriers and, to a lesser degree, realia which may also be part of archival collections. However, things such as submarines, aeroplanes or buildings certainly do not fall under this pattern, nor do works of art which belong to museums. Three principal types of inquiries are covered by the pattern shown in Figure 73:

1. *The existence and location* of things such as the particular call number of a file, the current holder of a collection of documents, or the availability of a finding aid.
2. Various *qualities* of things such as the range of a series, the title of a document, the type of a document, or the physical and intellectual structure of things.
3. The internal and external *authenticity* of things such as whether a document is the original or a reproduction.

As discussed in the previous chapter (IV:2.2), the archive and its collections consist of a physical and logical hierarchical structure. In the context of this study, the principal approach to representing the hierarchy mirrors to a large extent the proposal by Bountouri and Gergatsoulis (2011) who mapped EAD to the CRM.

The *physical* structure of an archival collection can be represented by instances of *E24 Physical Man-Made Thing*. Each instance of E24 represents an aggregation of physical items or, on the lowest level, potentially a single item, from the collection on a certain level of the hierarchy. The physical relationships can be modelled using the property *P46 is composed of (forms part of)*. The archival collection itself is represented by an instance of the class *E78 Collection*.

The physical collection is described (*P70 documents (is documented by)*) by an instance of the class *C7<sub>akm</sub> Finding Aid* which is a new sub-class of *E31 Document*. This class may represent any kind of archival aid describing any kind of archival collection (E24). The mapping of EAD by Bountouri and Gergatsoulis (2011) provides an example of how such a *C7<sub>akm</sub> Finding Aid* could be designed.

The *logical* structure of the archival collection, such as series, files or items, is represented by instances of *E73 Information Object*. This conceptual hierarchy can be modelled using the property *P46 is composed of (forms part of)* which connects, for example, a series (E73 with E55 “series”) with the files (E73 with E55 “files”) it contains.

As discussed in the context of the *Provenance* pattern (V:2.1.1), instances of *E73 Information Object*, for example the intellectual contents of a report, are carried (P128) by instances of *E24 Physical Man-Made Thing*, such as paper sheets.

Bountouri and Gergatsoulis (2011) also include *E33 Linguistic Object* in their mapping, which is not a relevant category in the context of the AKM since no questions specifically inquire about particular linguistic features of archival material.

The class *E55 Type* and the property *P2 has type (is type of)* can be used to provide the various kinds of *type* information (V:1.2) on the physical and conceptual archival materials. These could be the *type of form* such as photo or microfiche, the *type of function* such as cadre file, minutes or report, or the *type of collection*, such as personal estate but also the various logical units of the archival collection such as series or files.

The class *E35 Title* is used along with the property *P102 has title (is title of)* to provide a title for a work such as the texts and documents or pictures within an archival collection.

The class *E54 Dimension* along with the property *P43 has dimension (is dimension of)* covers the physical extent of the archival material.

As discussed in the context of the *Provenance* pattern (V:2.1.1), several questions generally inquire about “personal materials” or “personal documents”. These questions aim at finding any items known to have been either created, produced or kept by an actor over a certain time-frame or at the time of death. In an archive, such items can typically be found in personal estates, and personal materials and documents may also exist in other collections within the archival holdings.

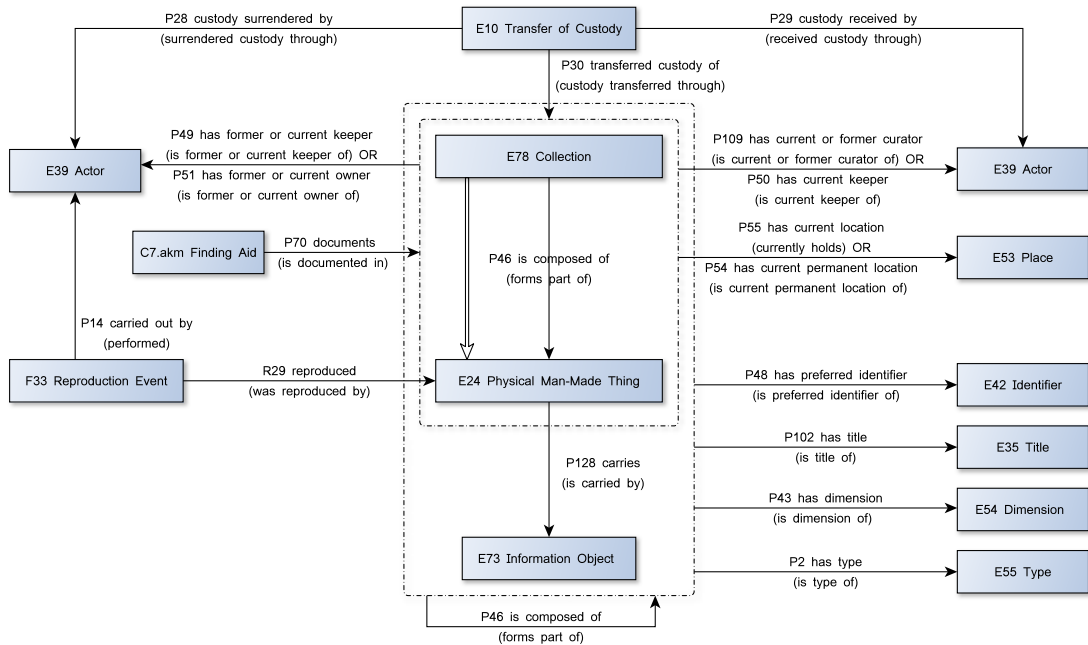


Figure 73 – General Documents pattern.

Instances of the class *E24 Physical Man-Made Thing* carry (P128) instances of the class *E73 Information Object*. In the context of the pattern *Provenance (V:2.1.1)*, the relationship to the actual “contents” of a physical thing such as a letter or photograph have already been discussed. In the context of the archival holding, this relationship may also designate the logical unit of the archival unit in the descriptive tree.

To a certain extent, the *Documents* pattern constitutes a connection between the past historical reality and the current archival reality in relation to the things (E24) produced and information (E73) created. This connection is illustrated by the class *E10 Transfer of Custody* which represents chains of custody and the transfer of custody from one actor to another. In the context of this study, the chain of custody ultimately leads to the archive as the current curator or keeper of physical things (E24), now archival materials. Particular examples of transfer of custody activities not only include the acquisition of material by an archive but also any kind of change in custody of items, such as acts of selling, theft, finding, or other kinds of voluntary or involuntary commission.

The property *P49 has former or current keeper (is former or current keeper of)* is a shorthand for the longer path through the event *E10 Transfer of Custody*: Instances of *E39 Actor* acquire physical custody (*P29 custody received by (received custody through)*) of instances of *E24 Physical Man-Made Thing* or *E78 Collection* through a *E10 Transfer of Custody* activity which renders physical custody (*P30 transferred custody of (custody transferred through)*) of an instance of *E24 Physical Man-Made Thing* to the *E39 Actor*.

Instances of *E39 Actor* do not necessarily need to have surrendered, legally or deliberately, custody to another instance of *E39 Actor*. For example, theft or robbery but also finding (*E24*) fall under *E10 Transfer of Custody*. However, the fact that a thing has been surrendered by a particular actor can be represented using the property *P28 custody surrendered by (surrendered custody through)*. A typical example is the transfer of custody of archival collections from one archive to another, or the routine acquisition of documents from state institutions.

The rules for the adequate description and identification of current archival units in the archive are covered by archival documentation principles and standards (*IV:2.2*). This study does not include within its scope the detailed representation of these principles in the AKM which, again, covers the recognizable interest of inquiries of archival users. The actual identification of a physical item in the archive is not explicitly covered, therefore, by any of the general patterns. The class *E42 Identifier* may hold a call number or any other means of identification necessary to retrieve an archival item.

The class *F33 Reproduction Event* is one of the very few entities that have been added to the AKM without explicit empirical evidence for its relevance. As a sub-class of *E12 Production*, instances of *F33 Reproduction Event* produce additional instances of *E24 Physical Man-Made Thing* for each reproduction, each carrying an instance of *E73 Information Object*. The intended purpose of this entity is to explicitly model whether or not a document has been reproduced. Examples of reproduction activities not only include archival reproduction, for example, the creation of digital copies or microfiche, but also any other reproduction conducted by other actors prior to the acquisition of a document by the archive.

The *Documents* pattern addresses questions related to archival material in the immediate and wider contexts of the archive. This refers to the existence and specific location of documents within the physical and conceptual structure of archival holdings, including questions related to transfer of custody and provenance in the widest sense prior to the current custodian. In contrast to resource discovery questions, these explicitly inquire about specific aspects of the existence or location and do not simply imply a need to learn about the current location of potentially relevant archival material.

### **Example - “Transfer of custody”**

Q039-07-03<sub>barch</sub> “Can you tell me whether the documents of each of the district committees of the Kulturbund were ceded coherently to the respective municipal archives after 1989?”

The general type of the question is *fact-finding*. The *wanted* entity is *<document.provenance>*. The relevant *given* entities are *type of group* (“district committees [of the Kulturbund]”, “municipal archives”), *type of activity* (“ceding [of documents]”), and *time-span* (“after 1989”).

The question is a *material fact* question since it inquires about the existence of a transfer of custody activity. The *provenance context* is *provenance [type of group]* since in the context of the current question a *type of group*, the district committees of the Kulturbund have ceded custody.

The *aboutness context* is *documents [provenance]* since the question specifically inquires after the provenance of a collection of documents.

Figure 74 shows a possible query pattern centred around a transfer of custody activity (E10) between the group of the district committees of the Kulturbund and the group of the respective municipal archives.

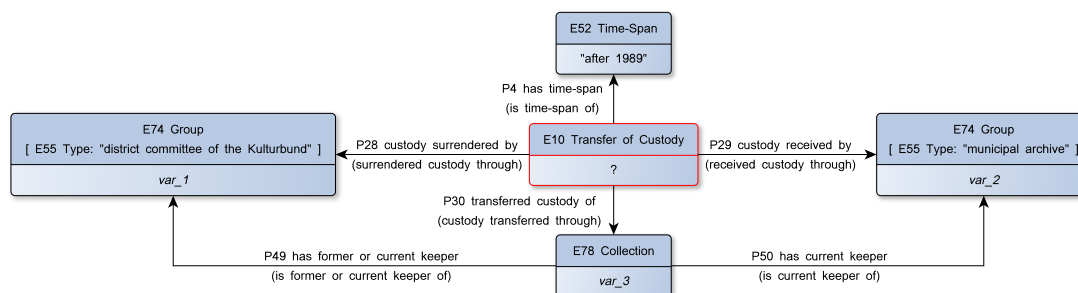


Figure 74 – Q039-07-03<sub>barch</sub>: Transfer of custody between two institutions.

The primary target of a query would be the identification of particular transfer of custody activities, surrendering the custody of any collection from a district committee of the Kulturbund to a municipal archive. The time-span information “after 1989” may not be considered crucial in this context but helps to further limit search requests.

Additional implicit and background knowledge may supplement queries; for example, about the geographical location and various known time-spans relevant to the existence of the archives involved. The initial question may also entail various follow-up questions such as the particular archives ceding and receiving documents, the dates of the transfer of custody activities, the nature and extent of the transferred collections and, of course, the location of these documents, which would result in a specific item question for particular collections.

### Example - “Extent of files”

Context                    “I found a possible access point (...) regarding the specific political work of the Bavarian delegates: 'Reichstag des Deutschen Reiches 1897-1938 (3487): Kommissions-(Ausschuss-) Akten 1868-1935 (521)'.”

Q052-01-02<sub>barch</sub>    “What is the extent of the holding?”

The general question type is *fact-finding*. The *wanted* entity is <document.extent>. The relevant *given* entities are *particular document* (“Reichstag des Deutschen Reiches 1897-1938 (3487): Kommissions-(Ausschuss-) Akten 1868-1935 (521)”), and *identifier*.

The question is a *material fact* question since it inquires about an observable material fact. The *provenance context* is *provenance [actor]*. The *aboutness context* is *documents [extent]*.

The user is referring to various documents previously requested in the same inquiry. The query pattern shown in Figure 75 would therefore have to be instantiated for every requested document.

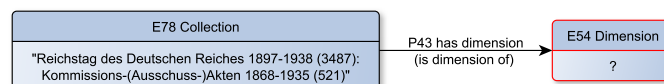


Figure 75 – Q052-01-02<sub>barch</sub>: The extent of files.

The question is a simple and basic one and could be answered by most existing archival information systems as long as the extents of particular units have been recorded in the archival aid.

### Example - “Containment of documents”

Q003-05-09<sub>barch</sub> “Do separate cadre files [of the KPD] exist for the following women or have their documents been filed with the files of their husbands? [person name wife] (born [maiden name], date of birth and place of birth unknown). (...) [person name husband] (...).”

The general question type is *resource discovery* <specific type>. The wanted entity is <document.provenance>. The relevant given entities are *type of document* (“cadre file”), *particular person* (“[person name wife], [person name husband]”), and *particular group* (“KPD”).

The question is a *material fact* question since it inquires about an observable relationship between two documents. The *provenance context* is *documentation* [actor]. The *aboutness context* is *documents* [document relation].

The query pattern shown in Figure 76 represents an adequate context for answering the above question. The information that the cadre file of the wife forms part of the cadre file of the husband may be known and recorded. In this case, the question could be answered by only querying the archival information system.

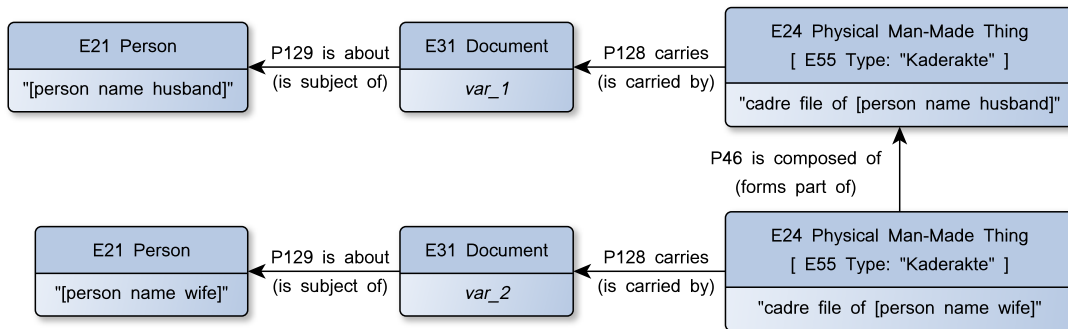


Figure 76 – Q003-05-09<sub>barch</sub>: Containment of documents.

It is more likely, however, that the cadre files of the husbands would have to be ordered and checked by the researcher if they indeed contain the cadre files of the wives. The documentation context has been omitted from the figure but could provide an indirect approach to answering the question.

**Statistics** The primary entities of interest are conceptualized slightly differently (V:1.5) than in the case of the other general patterns pertaining to the aboutness context. In the complete sample, the *Documents* pattern has been assigned to 34 questions representing 7.1% of all inquiries. Of these 34 questions 26 belong to the BArch sample, representing 7.5% of all questions in that sample, while five questions belong to the NAN sample and represent 6.1% of all inquiries in that sample. The primary entities of interest are distributed quite evenly compared to the other general patterns: In the whole sample ten questions representing 29% of all Documents questions are related to content information of which five belong to the BArch sample (19%) and five to the NAN sample (63%). *Document relations* with four and *extent* with five occurrences appear only in the BArch sample which represents 12% and 15% of the *Documents* pattern respectively. The primary entity of interest identity occurs six times (18%) in the context of the *Documents* pattern, of which five (19%) belong to the BArch sample and one (13%) to the NAN sample. Finally, nine *Documents* patterns relate to provenance, representing 26% of the *Documents* pattern, of which seven (27%) appear in the BArch sample and two (25%) in the NAN sample.

<i>Aboutness Context</i>	<i>All (n=467)</i>		<i>BArch (n=345)</i>		<i>NAN (n=131)</i>	
<b>Documents</b>	<b>34</b>	<b>7.1%</b>	26	7.5%	8	6.1%
- Documents [ content information ]	10	29%	5	19%	5	63%
- Documents [ document relation ]	4	12%	4	15%	0	-
- Documents [ extent ]	5	15%	5	19%	0	-
- Documents [ identity ]	6	18%	5	19%	1	13%
- Documents [ provenance ]	9	26%	7	27%	2	25%

Table 15 – Occurrences of the general pattern *Aboutness*.

**Summary** The *Documents* pattern covers the historical as well as archival context of information objects. In this regard, the subject matter of the pattern is slightly different and broader compared to the other general patterns in that it covers the transitional intersection between the historical reality in which information objects and their physical carriers participate, have been created and produced, as well as their current whereabouts within the archive. The *Documents* pattern mostly covers questions about qualities of things within the archive such as the extent of files or their history, such as transfer of custody activities.

The *Documents* pattern represents only those entities and relationships found in this particular sample. The pattern does not attempt to include archival context information which has not been identified as relevant to the sample.

Finally, the new class *C7<sub>akm</sub> Finding Aid* is introduced in order to distinguish the specific type of descriptive documentation of archives from other, more general types of documentation. The class represents any kind of archival aid (IV:2.2) specifically designed to provide information and access to archival holdings.

### 2.3 Summary

The summary for this section will focus on statistics and their interpretation within the context of the particular sample analyzed. The general patterns previously presented will be recapitulated only briefly, and the next section will provide a more comprehensive discussion of the general patterns.

The identified general patterns each cover a specific segment of a prototypical past reality, which can be distinguished into a context of provenance where patterns describe with different levels of detail how things have been generally created, produced or kept, and a context of aboutness where patterns describe to which specific historical contexts things refer. The *provenance context* (V:2.1) is represented by the general patterns *Provenance* (V:2.1.1), *Documentation* (V:2.1.2), *Self-Documentation* (V:2.1.3), and *Correspondence* (V:2.1.4) while the *aboutness context* (V:2.2) is represented by the general patterns *Aboutness* (V:2.2.1), *Events* (V:2.2.2), *Plans* (V:2.2.3), *Actors* (V:2.2.4), *Things* (V:2.2.5) and *Documents* (V:2.2.6).



The ontology CIDOC CRM proved to be a generally adequate means for the ontological representation of the subject matter of the interests of the inquiries. Nevertheless, new classes and properties were introduced in order to represent specific semantics. Furthermore, all inquiries in the sample were applicable to a general pattern. In other words, for each inquiry at least one appropriate query pattern could be formulated. This is not to say that, in some cases, the query patterns remained rather inclusive in terms of potentially retrievable information objects. In such cases, queries would have to be further specified by the user if more exact results were required.

Statistics have already been provided for each general pattern. Figure 77 summarizes the distribution of the general patterns pertaining to the *provenance context*. The general pattern *Provenance* is the most important in both samples with 156 occurrences (45%) in the BArch sample and 70 in the NAN sample (53%). In the whole sample, the *Provenance* pattern amounts to nearly half of all assigned general patterns from the *provenance context* with 226 occurrences (47%).

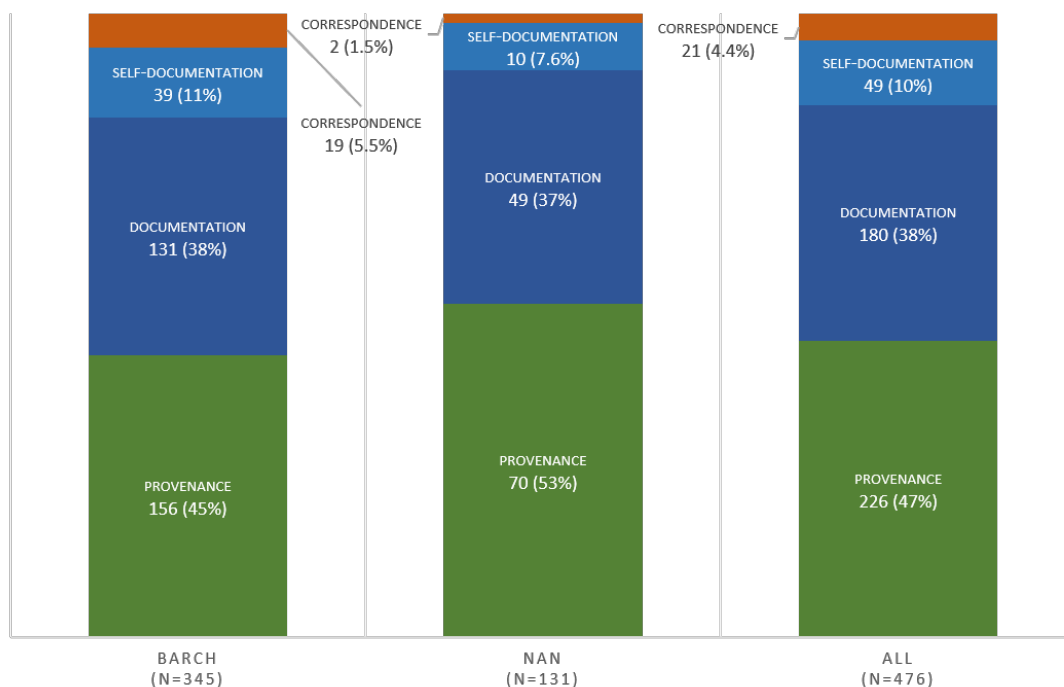


Figure 77 – General patterns belonging to the provenance context.

The general pattern *Documentation* is the second most frequent assignment in both samples with nearly identical percentages. In the whole sample, the *Documentation* pattern also comes second after the *Provenance* pattern with 180 occurrences (38%). However, since the general

pattern *Self-Documentation* is a specialization of the *Documentation* pattern, one may argue that the *Documentation* pattern is even the most important if both general patterns are counted together. As a distinct general pattern, *Self-Documentation* still amounts to 39 counts (11%) in the BArch sample and 10 in the NAN sample (7.6%). In the whole sample, the *Self-Documentation* pattern has 49 counts (10%). Finally, the general pattern *Correspondence* occurs more often in the BArch sample with 19 assignments (5.5%) compared to only 2 in the NAN sample (1.5%). In the whole sample, the *Correspondence* pattern amounts to 21 (4.4%) matches.

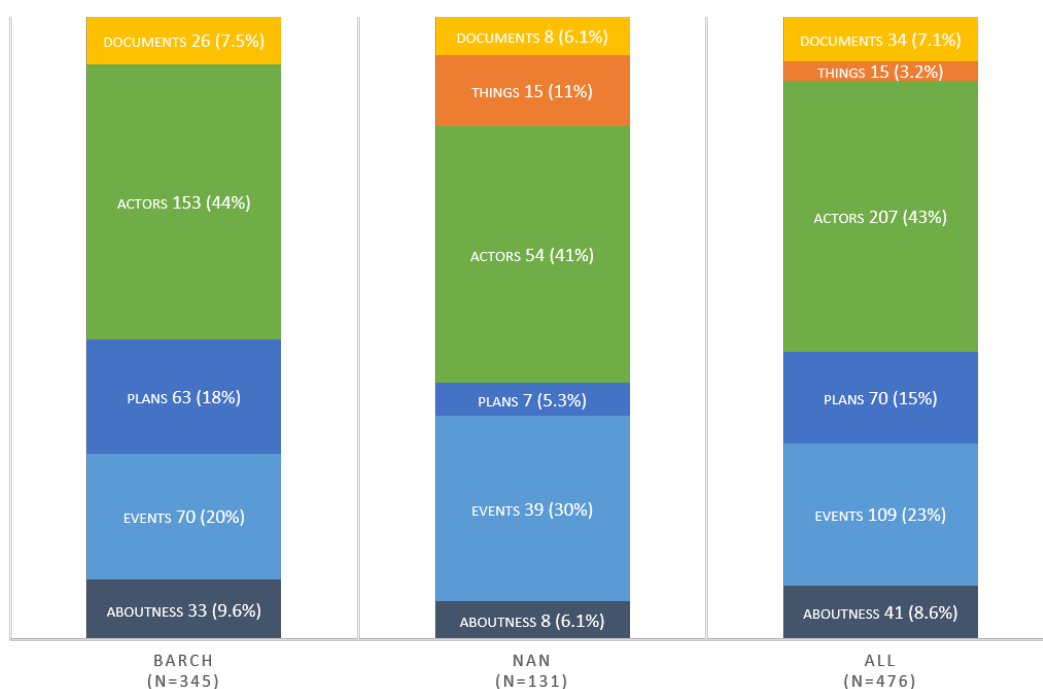


Figure 78 – General patterns belonging to the aboutness context per sample.

Figure 78 then shows the distribution of general patterns belonging to the *aboutness context*. The general pattern *Things* only occurs in the NAN sample with 15 (11%) counts. Additionally, the general pattern *Plans* is more frequent in the BArch sample with 63 occurrences (18%) as compared to the NAN sample with 7 occurrences (5.3%). Instead, the general pattern *Events* appears more often in the NAN sample with 39 assignments (30%) and only 70 assignments (20%) in the BArch sample. Overall, the general pattern *Actors* is the most important, amounting to 153 in the BArch sample (44%) and 54 in the NAN sample (41%). Finally, the general pattern *Documents* has a very similar frequency in both samples with 26 counts in the BArch sample (7.5%) and 8 counts in the NAN sample (6.1%).

Figure 79 then takes a look at the relative distribution of the general patterns among the two major categories of question types: resource discovery and fact-finding. The general pattern *Actors* constitutes the largest group in both cases with 143 occurrences (39%) with resource discovery questions and 64 occurrences with fact-finding questions (57%). The high percentage in the case of fact-finding is due to numerous questions asking for various facts about particular persons, as shown in Figure 82, most of which pertain to fact-finding.

Resource discovery questions further ask for information objects related to events and activities in 94 cases (26%), and related to plans in 54 cases (15%). The general pattern *Events* is used 15 times (13%) and *Plans* 16 times (14%) in the context of fact-finding. The general pattern *Things* is only relevant to resource discovery with 15 occurrences (4.1%), and the *Documents* pattern is relatively less important with 17 cases (4.7%) than in the context of fact-finding with 17 cases (15%).

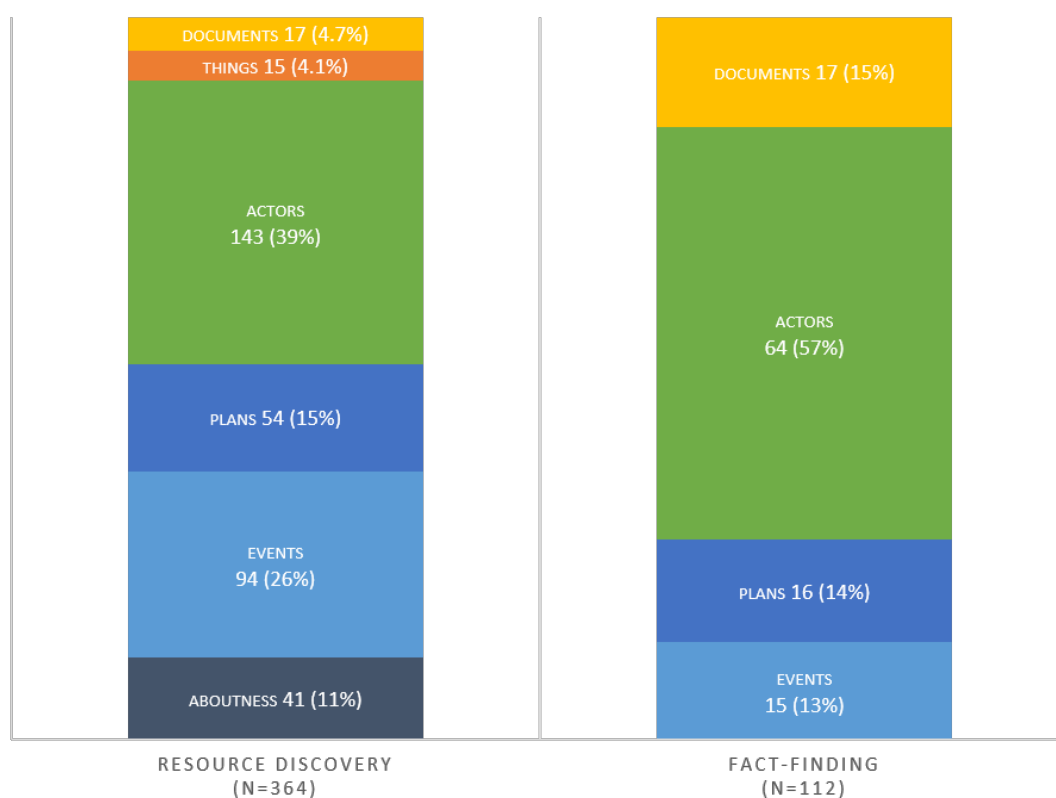


Figure 79 – General patterns belonging to the aboutness context per major type of question.

Since the general pattern *Aboutness* remains unspecific yet inclusive in terms of the subject matter of the information objects and aboutness of facts; that is, it is assigned only if these are

unknown or irrelevant, it is not assigned to any fact-finding questions. In the case of resource discovery questions, *Aboutness* appears 41 times (11%), mostly in the context of questions that request things by an actor where the provenance context is the decisive criterion. Once again, the aboutness context indicates the aboutness of information objects required in order to adequately meet the interest of inquiry, but it does not necessarily indicate the aboutness of the question per se, which is approximated by the wanted entities. For example, an unspecific request for documents created by a particular person is characterized by the wanted entity *actor.by* while the appropriate query pattern is realized by the general patterns *provenance [particular person]* and *aboutness [any]*, since the aboutness of the requested documents is undetermined.

While the particular distribution of the general patterns in the context of the sample already reveals a general tendency of the interests of inquiries, their specific combination promises to provide further insights.

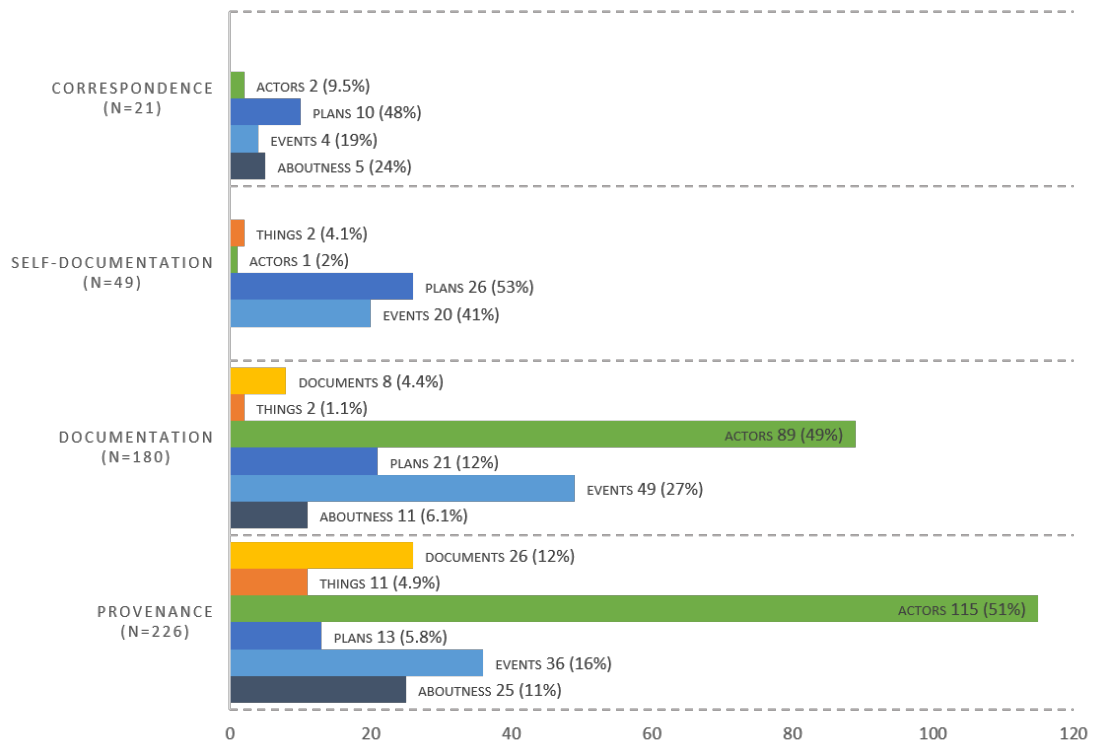


Figure 80 – Combinations of general patterns (percentages per general pattern from the provenance context).

Figure 80 lists the combinations of general patterns from both contexts. The importance of the general pattern *Actors* in the context of the *Provenance* and *Documentation* pattern is apparent:

in 115 cases (51%), the *Provenance* pattern occurs with the *Actors* pattern, while in the case of the *Documentation* pattern the *Actors* pattern occurs 89 times (49%). In the latter case, for example, this means that many questions are about documents as the product of the documentation of actors and their activities. In the case of the *Provenance* pattern, one tentative conclusion is that many questions generally inquire about actors where the provenance of documents (who created or kept these documents, is of no importance or unknown. Similar conclusions can be drawn for the general patterns *Events* and *Plans* in the context of the *Provenance* and *Documentation* patterns, even though they are, in both cases, less important than the *Actors* pattern.

The *Aboutness* pattern is assigned mostly when the *provenance* context is the decisive criterion; for example, things of or by a particular actor, or generally things resulting from documentation activities such as reports or minutes where the particular aboutness of these documents is unknown or irrelevant.

The *Self-Documentation* pattern is dominated by the *Plans* and *Events* patterns with 26 (53%) and 20 (41%) occurrences respectively. Most questions that inquire after things as the result of self-documentation are about either activities in general or aspects from the *Plans* pattern.

Finally, the *primary entities of interest* can be consulted in order to further explore the particular characteristics of the interests in the sample. A primary entity of interest has been assigned to each general pattern. For both general patterns from the provenance context and from the aboutness context, the identified primary entities of interest have already been introduced above (V:1.5). These entities further indicate the aboutness of information objects required in order to adequately meet the interest of an inquiry through the specification of a query pattern.

While the given entities constitute the basic building blocks of the general patterns and reappear as such in the ontological representation, the relationship between wanted entities and the general patterns and their primary entities of interest is, however, only indirect. Both provide descriptive characterizations of the users' interests on different levels and within different contexts and stages of the analysis. The wanted entities indicate a principle aboutness of either a resource or fact while the primary entities of interests point out the most important instantiated entity within a general pattern. The wanted entities are absorbed in the general patterns which are their further ontological realization, and provide the ontological means to address them, as demonstrated in the query patterns.

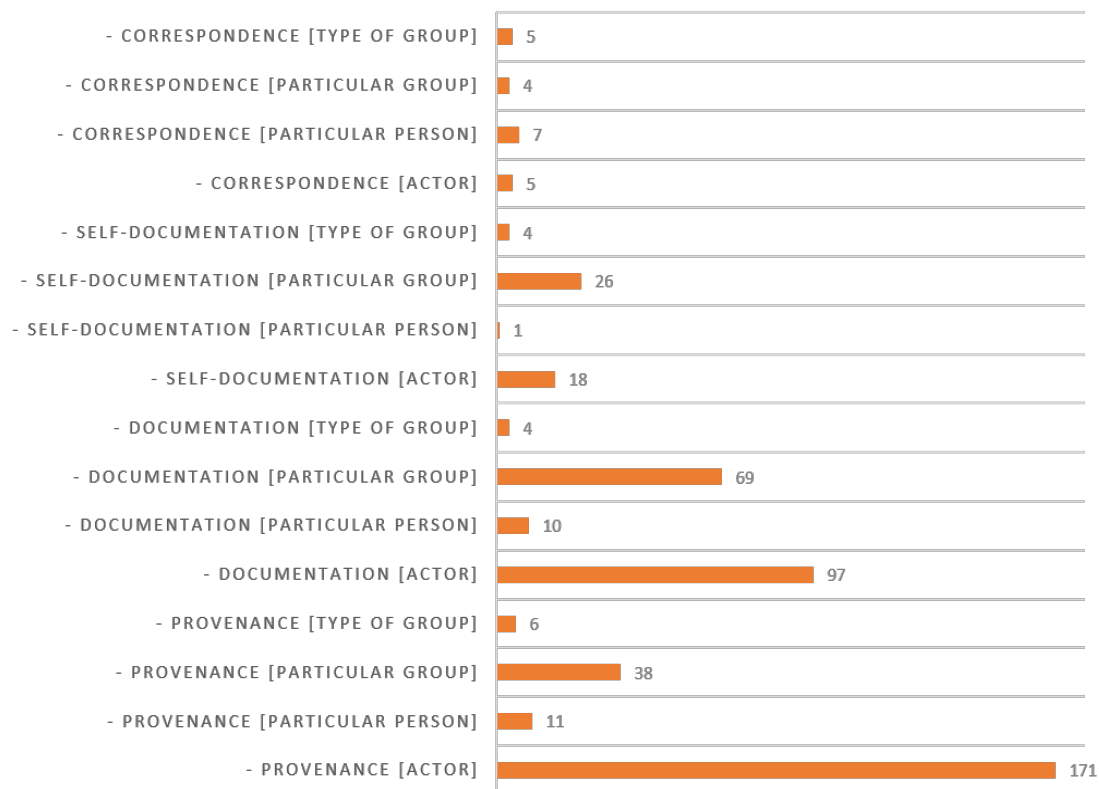


Figure 81 – General patterns of the provenance context with primary entities of interest.

Figure 81 shows the primary entities of interest per general pattern pertaining to the *provenance context*. The high amount of *[actor]* in the context of the general pattern *Provenance* as well as *Documentation* and *Self-Documentation* is apparent. While type of actors and particular actors; that is, persons and groups, occur 185 times amounting to 39% of all general patterns in the *provenance context*, *[actor]* occurs 291 times, which equals 61%. In the latter cases, the actor as the entity who has created, produced, or been the keeper of specific things is either not known or not recognizable in the inquiry, or the query patterns are meant to remain inclusive. This means that *[actor]* does not necessarily indicate a lack of information on the part of the user but may be intentional; the focus of the inquiry may be on the aboutness of the desired archival materials or facts. A prototypical request would be for any documents about a particular person, regardless of who created or was keeper of these documents. As Figure 80 indicates, a high percentage of *Provenance* as well as *Documentation* patterns indeed occur with the general pattern *Actors*. Furthermore, as shown in Figure 82, most general patterns *Actors* have a particular person as the primary entity of interest. On the other hand, as already mentioned, in nearly 39% of all cases, the inquiries inspire conclusions on more specific types or particulars of persons or groups.

Figure 82 shows the primary entities of interest per general pattern pertaining to the aboutness context. The general pattern *Aboutness* has not been assigned any primary entities of interest since the pattern corresponds to unknown or general aboutness of things or facts. For the general pattern *Documents* the primary entities of interest have been omitted in order to retain readability of the diagram.<sup>85</sup>

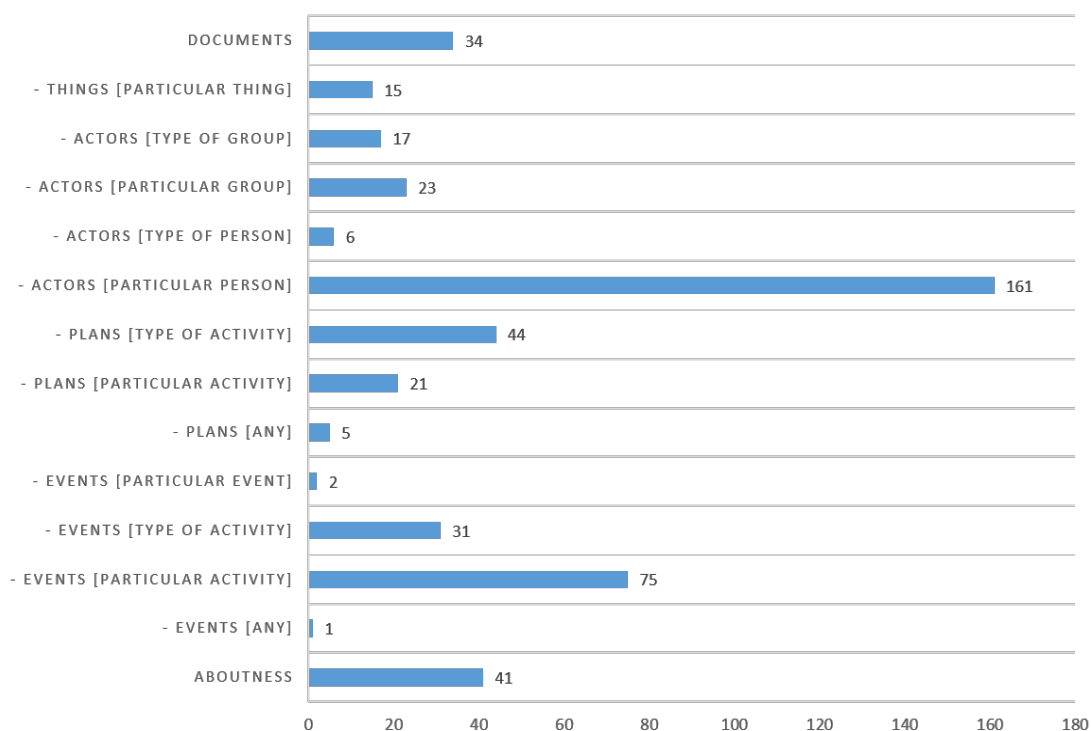


Figure 82 – General patterns of the aboutness context with primary entities of interest.

In the case of the general pattern *Things*, all information objects and facts were about *[particular things]*. For the general pattern *Actors*, most information objects and facts were also about *[particular persons]* or, to a lesser degree, *[particular groups]*. Similarly, in the case of the patterns *Plans* and *Events*, *[particular activity]* was clearly most important. Unspecific or unknown aboutness; that is, *[any]*, seldom occurs. Furthermore, types are also relatively unimportant, occurring in numbers only in the case of the general patterns *Plans* and *Events*.

The statistics provided in this section served to assess the relative importance of the various general patterns in relation to the sample of user inquiries. Furthermore, these have been interpreted in order to derive tentative insights into user needs when researching in archives,

<sup>85</sup> Statistics on their occurrences are provided in the context of the general pattern *Documents* (V:2.2.6).

based on the particular sample collected for this analysis. However, as mentioned earlier, the objective of this study is not to quantify and provide exact numbers on the relevancy of the interests represented by each general pattern, as several questions would fit more than one general pattern or overlap, as mentioned earlier.

In fact, the general patterns themselves are representations of user interests and embody user needs in archives. Each ontological entity in the context of a general pattern (their classes and relationships) constitute parts of these interests. Query patterns show how these entities can be used in order to formulate a query to fulfil an interest. The inquiries collected and analyzed have been instrumental in devising these general patterns as ontological representations of archival user needs. As such, the statistics and numbers only provide tentative indications of the relevancy of specific general patterns and entities within these patterns. In principle, all ontological entities in the general patterns are relevant and have been formulated based on the empirical evidence found in the sample. The conclusion (VII) will discuss future work regarding the further development of the AKM and its general patterns.

### 3 Summary: Archival Knowledge Model

The ontology *Archival Knowledge Model* (AKM) developed in the previous sections pertains to the subject matter of the interests of archival and historical inquiry as interpretatively identified in user inquiries submitted to archives. The AKM is based on the interpretative analysis of 476 single inquiries (III) extracted from user files collected from the German Federal Archives and the National Archives of Norway. The understanding of the users' epistemology as delineated in the previous chapter (IV:2) provided the criteria from which to assess the relevance of the kinds of facts that deserve modelling in the ontology; that is, those of primary interest to the users.

The AKM is the set of common and abstract entities which are a formal and explicit representation of the various interests interpretatively identified in written natural language inquiries submitted to archives. As an ontology, the AKM formally and explicitly represents the subject matter to which the interest of archival users and their inquiries pertain. The AKM succeeds in reducing complexity of phenomena and interests to a relatively small set of entities. Each general pattern constitutes an aggregation of classes and properties, thus representing specific prototypical segments of this subject matter. Together, the general patterns constitute the AKM. Query patterns are exemplary and selective instantiations of two general patterns demonstrating adequate potential queries against hypothetical archival information systems.

Furthermore, based on the material facts described by the AKM, we can conclude which kinds of information and contexts should be made explicit and encoded in an archival knowledge base. Consequently, the AKM can be used to analyze existing and devise new archival data standards and to conceptualize access and query facilities for archival knowledge bases and information systems. As such, the AKM is to be understood as a supplement and additional, non-archival descriptive access point to existing archival descriptions and information systems.



Even though the AKM contains all classes and properties that, together, allow us to describe and represent the subject matter of the interests to which the archival inquiries pertain, each general pattern addresses typical and prominent segments of this subject matter and must therefore be discussed separately. For more extensive and detailed accounts as well as definitions of newly created extensions of the CRM, see the respective section of each general pattern and the appendix. The following provides a concise summary of all ten general patterns.

- Provenance (V:2.1.1)
  - The *Provenance* pattern represents on a generic level three notions of provenance of intellectual and physical things. The first two are the *intellectual creation* and the *physical production* of things by an actor, such as conceiving of a text and then writing it down on a sheet of paper. These two acts are typically not distinguishable from one another. The third notion of provenance refers to the things *kept by* an actor which does not entail legal ownership but only possession of the thing during a certain period of time. All three notions combined cover the idea of any things *of* and *by* an actor, including the notion of personal things of an actor. The *Provenance* pattern suggests that the explicit representation of such basic contexts of creation, production and possession is highly relevant to users supporting generic yet inclusive questions: Who created specific things, and who produced these things? Who has kept these things at least at some point in the past? The pattern does not further qualify these contexts other than by specifying the temporal and geographical extent of these activities. The general patterns Documentation, Self-Documentation, and Correspondence further elaborate these contexts of provenance.
- Documentation (V:2.1.2)
  - The *Documentation* pattern covers acts of direct or indirect *observation* or *documentation* of events or the activities of actors, which typically result in documents about the observed. The pattern differentiates between the act of documentation and the documented act itself – which do not necessarily occur at the same time – and thus distinguishes explicitly between the actor carrying out the former activity and the actor carrying out or participating in the latter activity. Furthermore, the pattern represents the purposeful conduct of documentation acts based on a *mandate*, which covers any form of official or unofficial assignment to carrying out observational acts such as taking minutes or surveillance. The mandate itself represents a pivotal concept for inquiry within the archival but also the historical context, which is essential to identifying potential creators or keepers of relevant documents. The *Documentation* pattern thus covers essential and fundamental questions of historical and archival inquiry: Who has or must have written about an event or activity? Who has or must have written about an actor? Who had or must have had the mandate to conduct a documentation activity?

- Self-Documentation (V:2.1.3)
  - The *Self-Documentation* pattern extends the *Documentation* pattern and specifically describes the direct or indirect documentation of one’s own activities; that is, *self-observation*. An actor presents the perspective on his or her own activities that is intended for external consumption. Such documents may be official or unofficial, internal or public; however, the pattern specifically emphasizes internal accounts such as minutes or accountability reports, created by state and administrative agencies. Their value for historical research is that they result from routine and the daily business of administrative bodies and other agencies, and document their plans, decisions and discussions from internal and potentially from “unofficial” perspectives. The *Self-Documentation* pattern thus in particular covers questions regarding the internal and inner dispositions of actors: How did an actor arrive at a result or decision? What were the reasons, motives or plans of an actor? What did an actor want others to believe about their actions? As shown in the previous section, the *Self-Documentation* pattern is likely to occur with the *Plans* pattern.
  
- Correspondence (V:2.1.4)
  - The *Correspondence* pattern represents three basic aspects of oral or written communication acts between two or more actors involving information objects. The first two notions address the intentional or unintentional *designation* of an information object to an actor – as, for example, propagated during a speech – and the phenomenological knowledge regarding the *addressee* of an information object as, for example, given in the address line of a communiqué. The third notion resembles the idea of *taking notice* of information by having kept and thus received in some form an information object. The *Correspondence* pattern thus addresses a range of questions regarding recorded acts of communications: Who was the designated audience of an information object? Who was the intended addressee of an information object? Who has or must have taken notice of an information object? Who had or must have had knowledge of a piece of information?
  
- Aboutness (V:2.2.1)
  - The *Aboutness* pattern covers the basic relationships of general reference and primary subject or topic between an information object and another entity. The focus rests on basic topical questions and inquiries in which the subject relationship cannot be specified because it is too broad, unknown, or irrelevant, or is intended to be inclusive. The pattern thus covers generic yet inclusive questions regarding the general aboutness of an information object: Of which entity does an information object speak? Which entity is referenced by an information object? The *Aboutness*

pattern does not further qualify the historical context of the entities referred to other than by specifying their temporal and geographical embedding. The general patterns *Events*, *Plans*, *Actors*, *Things* and *Documents* each further elaborate the context of aboutness.

- Events (V:2.2.2)
  - The *Events* pattern covers the historical context of unintentional events and intentional activities, in which actors actively or passively participated, and the acts carried out by actors. The pattern allows for the representation of events and activities within their temporal and spatial contexts and of who was actively or passively present during their occurrence. Complex events and activities are represented by fundamental relationships of influence, sequence, and part-whole. The pattern therefore covers questions regarding things providing information or evidence about events or activities: Did an event or activity occur? When and where did an event or activity occur? How did an event or activity occur? What happened to an actor? Who carried out an activity? The *Events* pattern allows for the assertion of basic motivational or influential relationships between events and activities. The *Plans* pattern, however, further covers details of the planning and wanting of activities.
- Plans (V:2.2.3)
  - The *Plans* pattern extends the *Events* pattern (V:2.2.2) by the two basic notions of *planning* and *wanting* of activities by actors. The pattern specifically allows for descriptions of the planning of activities and their plan-based execution, and more specifically the discussion and creation, the deciding and abandoning of plans, as well as the beginning and ending of a will to execute a plan. The pattern particularly distinguishes between the plan itself, the expression of a will to act, and the document which may contain a plan and expression of will. This distinction allows for a general reference to plans and enactments of plans, and in particular questions related to the intentions and aims of actors, to what extent plans have been applied and executed, and who wanted plans. This includes intentions to execute plans versus what is known or has been recorded regarding the actual execution and enforcement of plans. The *Plans* pattern thus introduces a temporal descriptive perspective looking into the future from any given point in the past, recorded in an information object.
- Actors (V:2.2.4)
  - The *Actors* pattern covers actors, persons and groups, their lives and history. The pattern can be understood, similar to the *Plans* pattern, as another extension of the *Events* pattern, in this case focusing on the particular qualities and contexts of those who carry out activities and participate in events. The *Actors* pattern can thus be

considered biographical, covering the typical activities in the lives of actors such as the joining or leaving of various kinds of groups, including marriage, birth and death, and the relationships and encounters between individuals and groups. As seen in the previous section, the *Actors* pattern mostly focuses on particular persons and groups and occurs frequently in the context of fact-finding questions. The *Actors* pattern therefore covers general or specific and often factual biographical questions: When and where did an actor come into existence and when was an actor taken out of existence? What was the relationship of an actor with another actor? When did an actor live and operate? Where was an actor at a specific time? What did an actor do?

- Things (V:2.2.5)
  - The *Things* pattern covers a basic historical context surrounding the history of physical man-made things. The empirical evidence supporting the pattern is not strong in the sample of this study, which is not surprising since historical inquiry as conceptualized here, and as might be expected in the context of state archives, does not frequently include object history. However, the pattern covers an existent segment of historical inquiry focusing on the history of objects such as submarines or buildings: Who produced the object? Who kept the object? In which event or activity was object involved? Where and when was the object involved in an activity or event?
- Documents (V:2.2.6)
  - The *Documents* pattern covers the historical context as well as a rudimentary archival context of information objects, at least to the extent that it has been pertinent in this sample. In this regard, the subject matter of the pattern is slightly different and broader than the other general patterns in that it covers the transitional intersection between the historical reality in which information objects as well as their physical carriers participated and were created and produced, and their current whereabouts within the archive. The *Documents* pattern mostly covers questions about qualities of things within the archive, and their history such as transfer of custody, the physical or logical relationship to one another, or their extent. Only those entities and relationships found in the particular sample of this study have been included, and no attempt was made to include comprehensive archival context information.

This chapter has developed ten general patterns based on the interpretative analysis of written natural language inquiries submitted to archives and by the ontological representation of the subject matter to which the interest of the inquiries pertain. The query patterns have exemplified the relationship between the inquiries and the ontological model, the AKM. The next chapter will introduce four archival patterns, which exemplify the relationship between the ontological model, the AKM, and selected examples of real-life archival data.

## Chapter VI

# Application

In this chapter, four exemplary *archival patterns* (IV:1.2.2) will illustrate how real-life metadata relates to the *Archival Knowledge Model* (AKM). As has been indicated in Figure 8 (IV:4) the general patterns can be understood as formulas or schemas for query patterns and archival patterns. While the former represent the relationship between user inquiries and the ontological formalization of the primary subject matter of their interest, archival patterns exemplify the relationship between archival documentation and the AKM itself; that is, the formalized subject matter of the interest of users' inquiries and the implicitly or explicitly represented knowledge. Each example is meant to demonstrate selected aspects of the relationship between existing archival data structures and the AKM. Furthermore, the examples show the interpretative transition from the empirical archival data to the ontological level of the AKM; that is, the immediate relationship between particular *Encoded Archival Description* (EAD) elements and their contents, and the entities in the general patterns.

Three examples are provided by the Federal State Archives of Baden-Wuerttemberg<sup>86</sup> (*Landesarchiv Baden-Württemberg*), hereafter "LaBW", and one has been obtained from the German Federal Archives (BArch). Two examples cover files from the 20th century and two from the 19th century. All examples are excerpts taken from existing finding aids encoded with EAD, chosen because it is becoming a widely used international de-facto standard for encoding archival aids. For this reason, no examples have been taken from the National Archives of Norway (NAN) since they do not use EAD encoded finding aids. The examples obtained from the LaBW demonstrate that, even though the general patterns have been developed based on inquiries collected from the BArch and the NAN, the archival patterns nevertheless also apply to data from other archives.

The previous chapter describes how information in written natural language inquiries can be interpreted and represented in terms of a formal data structure using CRM concepts. The goal of this study is not to describe or assume an algorithm that automatically converts free text, either from user inquiries or from archival data, into a formal data structure such as an RDFS/OWL implementation of the CRM.<sup>87</sup> Rather, the objective here is to motivate and demonstrate on an exemplary basis that existing archival descriptions potentially carry both explicit and implicit information pertaining to the general patterns of the AKM, and that archivists are in a position to encode equivalents of the information in these texts in such a formal data structure. The examples indicate that, while archivists document rich information, they generally neither structure nor link this information sufficiently or adequately.

---

<sup>86</sup> <http://www.landesarchiv-bw.de/web/>

<sup>87</sup> <http://erlangen-crm.org/>

# 1 Encoded Archival Description

*Encoded Archival Description*<sup>88</sup> (EAD) permits the encoding of information about the origins and emergence of the archive and its records, and provides information about its contents and scope. The EAD standard is a XML mark-up language which provides the flexibility necessary to accommodate the various nuances in archival documentation and description practices. As such, EAD can be used to encode archival aids which describe one or more archival collections, themselves consisting of many files or items organized according to provenance in sequential order and in a contextualizing hierarchy. In recent years, the EAD has indeed become a widely accepted de facto standard for the encoding of digital archival aids, even though its implementation in different systems exhibits inconsistencies in terms of the use of data elements and labelling terminology (Kim, 2004).

The EAD standard is the result of a by now international effort to standardize the encoding of archival descriptions (Frusciano, 2008a). One primary reason for undertaking a standardization of the encoding of archival descriptions was, according to Pitti, that archivists needed assurance that the valuable information they had curated would remain usable in the future, despite technology changes. Only hardware- and software-independent encoding of that information would provide this assurance to a reasonable extent. Furthermore, in order to support sophisticated indexing, navigation and display of archival information, and thus also thorough and accurate access and control of archival descriptions, their logical components and relationships would need to be accurately identified and encoded in machine-readable form. Uniform and standardized descriptions and encoding of archival information would elevate the “present chaos of irregularities” in archival descriptions and facilitate the identification and comprehension of the latter by archivists and users of archives alike. Standardization of encoding and description was understood to contribute to the realization of “the long-cherished dream” of enabling virtual universal, union access to archival materials, independent of the physical location or dispersion of the holding (Pitti, 1999).

The general outline of any EAD document comprises the root element <ead> and two basic general sections marked by the elements <eadheader> and <archdesc>. The <eadheader> element contains bibliographical and descriptive information to identify a finding aid document such as its title, author, creation date, or edition. The <archdesc> element then holds information about the archival collection as a whole and – within subsequent descendant <c> elements – information about classes, series, sub-series, files, and items represented in a hierarchical and sequential order.

As has been previously discussed (IV:2.2), files and, much less frequently, items generally constitute the smallest unit within the archival description. All other intermediate levels normally provide further context for a file. On each level in this descriptive tree constituted by

---

<sup>88</sup> The current version of EAD is the 2002 schema as found on <http://www.loc.gov/ead/>. For the following also see the EAD Tag Library at <http://www.loc.gov/ead/tgLib/index.html>.

<c> elements, various descriptive metadata fields hold information pertaining to each current level starting with and depending on the characteristics of the described holding with the whole archival fond and then proceeding down to the file or item level. Only few elements are declared mandatory by the EAD schema, in order to allow the necessary flexibility for accommodating the most particular descriptive practices. Several different EAD dialects exist, all of which are subsets of the full EAD model and specify particular restrictions on the use of certain elements and attributes such as the values of type attributes.

One of the most important EAD elements is <unittitle>, which holds a “name of the described materials”<sup>89</sup> at any level of the descriptive tree such as the title of the current holding or the name of a file. Other relevant examples of EAD elements accommodate a time-span during which the material has been created. This is recorded in <unitdate>, a call number for the archival material provided in <unitid>, or the name of the repository where the material is kept is found in <repository>. The contents of an archival unit may be given, for example, in <scopecontent> as a “prose statement summarizing the range and topical coverage of the described materials”<sup>90</sup> in order “to assist readers in evaluating the potential relevance of the materials to their research”<sup>91</sup>, or in <abstract> containing a “brief summary of the materials being described, used primarily to encode bits of biographical or historical information about the creator”<sup>92</sup>. EAD also allows for the encoding of “key terms and reference pointers”<sup>93</sup> within the element <index> providing, for example, person names or place names relevant to the materials described.

## 2 Archival Data and the AKM

All four examples feature one particular <c> section taken from a finding aid and centre around the <unittitle>, which is an essential element and can be expected in any archival finding aid encoded with EAD. The examples demonstrate how the information typically contained in these elements in combination with few additional entries in other elements already provide plentiful contextual knowledge representable with the general patterns of the AKM. The goal is not to show or decide on any “correct” representations but to demonstrate a logical correlation exists between the AKM and EAD.

---

<sup>89</sup> <http://www.loc.gov/ead/tglib/elements/unittitle.html>

<sup>90</sup> <http://www.loc.gov/ead/tglib/elements/scopecontent.html>

<sup>91</sup> <http://www.loc.gov/ead/tglib/elements/scopecontent.html>

<sup>92</sup> <http://www.loc.gov/ead/tglib/elements/abstract.html>

<sup>93</sup> <http://www.loc.gov/ead/tglib/elements/index-element.html>

### Example “Record of Interrogation”

The first example is taken from a finding aid created by the BArch and which has the unofficial name *Roter Koffer*<sup>94</sup> as it describes the contents of a red suitcase found within the estate of Erich Mielke and which belonged to Erich Honecker. The suitcase contained documents about a trial for high treason against Erich Honecker at the People’s Court (*Volksgerechtshof*) as well as reports and analyses created by the Ministry of State Security of the GDR (*Ministerium für Staatssicherheit*, MfS) at a later date about the interrogations related to the trial.

The descriptions found in this particular finding aid are quite informative and relatively detailed, partly due to the prominence and historical value of the contents of the suitcase. The XML snippet shown in Figure 83 represents one <c> section taken from that finding aid. It describes a file containing eight paper sheets pertaining to the interrogation of Sarah Fodorova on 9th November 1936.

```
<c04 level="file">
  <did>
    <physdesc>
      <extent>8 Bl.</extent>
      <physfacet type="damage">Risse im ersten Blatt</physfacet>
    </physdesc>
    <unitid type="bestellnummer">15</unitid>
    <unittitle>Vernehmungsprotokoll Sarah Fodorova vom 9. Nov. 1936</unittitle>
    <unitdate normal="1936-01-01/1936-12-31">1936</unitdate>
    <origination label="final">Oberreichsanwalt beim Volksgerechtshof: R 3017 ORA / VGH
      , 17 J 22/36, Bd.12</origination>
    <abstract>Enthält: Vernehmung von Erich Honecker nach Gegenüberstellung mit Sarah
      Fodorova.</abstract>
  </did>
  <processinfo><p>Die Blätter (Kohlepapierdurchschläge) tragen die Paginierung 108-115
    in Blaustift und stammen aus einer ursprünglich genähten Akte. Sie lagen ohne
    Umschlagdeckel im roten Koffer.</p></processinfo>
  <index>
    <indexentry>
      <persname>Fodorova, Sarah</persname>
    </indexentry>
    <indexentry>
      <persname>Honecker, Erich</persname>
    </indexentry>
  </index>
</c04>
```

Figure 83 – XML-Snippet from EAD file “Roter Koffer”

The <unittitle> element contains a range of implicit information. There was an interrogation (*Vernehmung*) of a person named Sarah Fodorova on the 9/11/1936, which was documented by minutes (*Vernehmungsprotokoll*). Such entries, which compress a range of historical facts, can be understood as examples of the wealth of hidden knowledge in metadata fields. The

<sup>94</sup> “Roter Koffer” translates to “Red Suitcase”. For background information on this holding see [http://www.bstu.bund.de/DE/Wissen/Aktenfunde/Roter-Koffer/roter-koffer\\_inhalt.html](http://www.bstu.bund.de/DE/Wissen/Aktenfunde/Roter-Koffer/roter-koffer_inhalt.html).



AKM permits decisions on which implicit information should be made explicit and how this knowledge should be structured.

Figure 84 shows an exemplary instantiation (of parts) of the general pattern *Documentation* (V:2.1.2) based on the information available in the <unittitle>. In this representation the information is explicit and formalized according to a pattern which is relevant to a broad range of information needs deriving from typical user inquiries and which focuses on the acts of documentation regarding the activities of others. In this case, the act documented is a statement made by a person called Sarah Fodorova. While the XML snippet does not mention who conducted the interrogation, we can assume that the archivist who created the description likely knew more about this, or was at least in a position to make reasonable assumptions, such as the People’s Court (*Volksgerichtshof*) as a group actor, and therefore would have been able to record this information.<sup>95</sup>

Note that the time-span “1936” provided in <unitdate> pertains to the documentation act and not necessarily to the interrogation. The activity that created the documents; that is, the paper sheets contained in the record, is not necessarily identical with the date of the interrogation. The documentation class represents the act of creating the files through documentation and aligns with the semantic of <unitdate>.

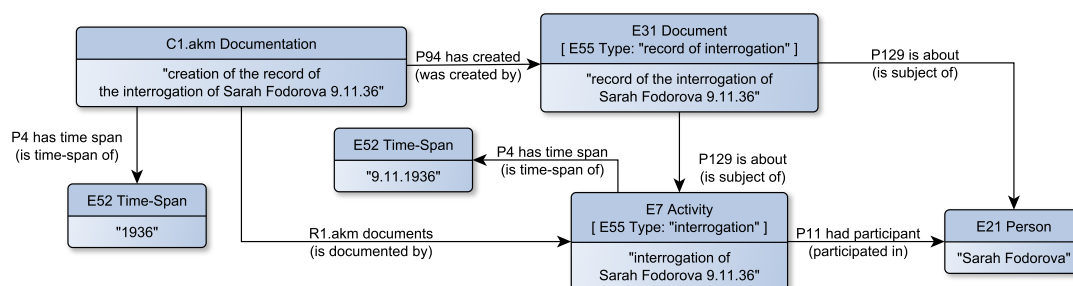


Figure 84 – The information from the <unittitle> represented explicitly (EAD file “Roter Koffer”).

*Self-documentation* is not used in this example since the interrogated person, Sarah Fodorova, is not herself conducting the documentation; that is, writing the report on the interrogation or taking notes in preparation for doing so. Furthermore, in the context of this example, we cannot assume that the statement made by Sarah Fodorova is done so on a voluntary basis; indeed, the report is unlikely to contain her own viewpoint but rather that of the interrogator. The expression of one’s own viewpoint is an important criterion for self-documentation. Choosing the *Documentation* pattern in order to represent the interrogative act does not preclude asserting

<sup>95</sup> Note that this information could be available at another level of the descriptive hierarchy within the finding aid from which the example has been taken.

that the interrogating actor is also conducting self-documentation as part of the interrogation. If it is not possible to ascertain clearly whether an event or activity fits a particular class then the more general class should be chosen. The example demonstrates that the acts of documentation or self-documentation may in this case have occurred close to the documented acts and might even have been temporarily congruent.

Finally, the intellectual task of the archivist when creating the title would remain the same when serving the seemingly more complex general patterns of the AKM. The “mechanical” effort might differ in that it is quick and easy to simply type in a literal text. However, this is a question of implementation and of proper tool design in the creation of archival aids. On the contrary, the archivist’s intellectual work is preserved here in a relevant and explicit representation while, in a plain and literal text, it would be lost and much less accessible.

### Example “Minutes of the Council of Ministers”

While the previous example was based on a documentation activity, the XML snippet shown in Figure 85 is a typical example of a self-documentation activity. The example is taken from a finding aid in the LaBW. The <unittitle> plainly states “Minutes of the Council of Ministers” (*Protokolle des Ministerrats*) and the <abstract> further specifies that the minutes pertain to its 1st - 35th meetings. The Council of Ministers in question is a body of the ministry of the German federal state Baden-Wuerttemberg as described at higher levels of the description in the finding aid. The <unitdate> further indicates that the documents were created between 25th April and 22nd December 1952.<sup>96</sup>

```
<c level="file" id="labw-1-783976">
  <did>
    <unitid>EA 1/105 Bü 1</unitid>
    <unittitle>Protokolle des Ministerrats</unittitle>
    <unitdate normal="1952-01-01/1952-12-31">25. April - 22. Dezember 1852</unitdate>
    <physdesc>
      <dimensions>Folio (Höhe x Breite)</dimensions>
    </physdesc>
    <abstract>1. - 35. Sitzung</abstract>
  </did>
  <otherfindaid>
    <extref xlink:role="url_archivalunit" xlink:href="http://www.landesarchiv-bw.de/
      plink/?f=1-783976">Titelaufnahme im Angebot des Archivs</extref>
  </otherfindaid>
</c>
```

Figure 85 – XML-Snippet from EAD file “labw-f-1-5717”

Figure 86 shows an archival pattern describing the “first meeting” of the Council of Ministers. The time-span provided in <unitdate> refers to the general time-frame during which the contained files were created. In the case of the current example, it is reasonable to assume that

<sup>96</sup> Note that the literal value in <unitdate> contains a typo; the proper date is 1952, not 1852.

the documented meetings of the Council of Ministers also fall into the same general time-frame.

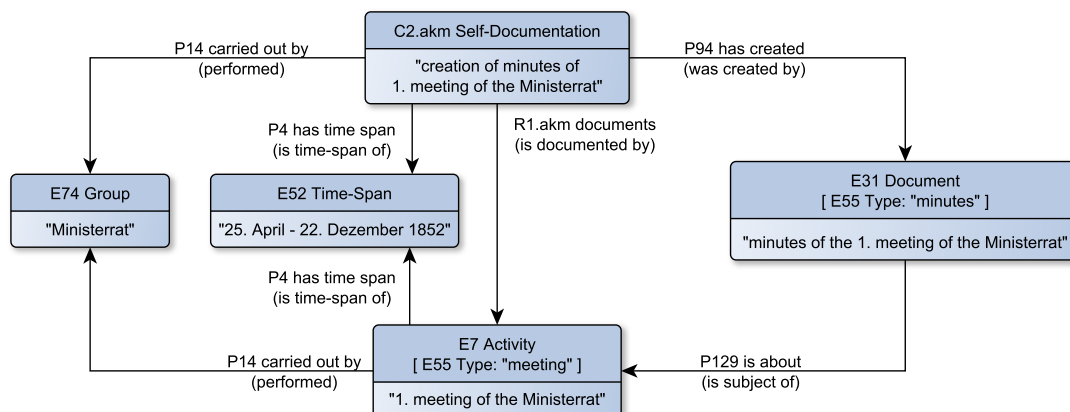


Figure 86 – The information from the <unittitle> represented explicitly (EAD file “labw-f-1-5717”).

Whether the meetings of the Council of Ministers should be represented as one activity extending over a longer time-span or as separate single activities depends on the intended descriptive scope and detail of a possible representation based on the AKM. Both variants are feasible and their implementation mostly depends on available resources and other technical prerequisites. However, if a particular explicit logical correlation exists within the data, in principle, all forms inferred can be automatically created by reasoners.

The Council of Ministers can be said to have self-documented its own activity since the minutes created reflect the Council’s own consensus and viewpoint. While the documents express neither the opinion of the stenographer nor of the person who took the notes and wrote up the final versions of the minutes, they do represent the collective opinion of the Council of Ministers of this particular meeting.

The example demonstrates the application of the general patterns *Self-Documentation* and *Events* and further shows that the provenance context may be more important, or at least more prominent, than the aboutness context. The aboutness here is simply an activity, the first meeting, while the provenance is described elaborately.

### Example “Petition of the ‘Geheime Rat von Edelsheim’”

The first two examples pertained to more recent times. The third example falls into an earlier era, the beginning of the 19th century, and utilizes primarily elements from the general patterns *Correspondence* and *Plans* as well as *Events*. The XML snippet shown in Figure 87 is taken from a finding aid of the LaBW and represents the description of one file. The file contains one written

document, a letter with a petition of the “Geheime Rat von Edelsheim” directed to “Karlsruhe” in order to obtain an excerpt from the knights’ cadastre regarding a debt claim from Freiherr Georg Wilhelm von Massenbach.

```
<c level="file" id="labw-1-603561">
  <did>
    <unitid>E 40/10 Bü 60</unitid>
    <unittitle>Gesuch des Geheimen Rats von Edelsheim in Karlsruhe um einen Auszug aus dem Ritterkataster des Kantons Kraichgau wegen einer Kapitalforderung an den Freiherrn Georg Wilhelm von Massenbach</unittitle>
    <unitdate normal="1822-09-11">11. September 1822</unitdate>
    <physdesc>
      <extent>1 Schr.</extent>
    </physdesc>
    <unitid type="Altsignatur">Ritterschaft 01 b/07 (Beilage)</unitid>
    <unitid type="Altsignatur">E 41 Verz. 64 Fasz. 16/07 (Beilage)</unitid>
  </did>
  <otherfindaid>
    <extref xlink:role="url_archivalunit" xlink:href="http://www.landesarchiv-bw.de/plink/?f=1-603561">Titelaufnahme im Angebot des Archivs</extref>
  </otherfindaid>
  <index>
    <indexentry>
      <persname>Edelsheim, Wilhelm Freiherr von; badischer Geheimer Rat, 1774-1840</persname>
    </indexentry>
    <indexentry>
      <persname>Massenbach, Georg Wilhelm Freiherr von</persname>
    </indexentry>
    <indexentry>
      <geogname>Kraichgau; Ritterkanton</geogname>
    </indexentry>
  </index>
</c>
```

Figure 87 – XML-snippet from EAD file “labw-f-1-603494”

The archival pattern shown in Figure 88 details the rich amount of implicit historical facts contained in the record description. The interpretation of the information contained in the EAD description is only preliminary. The claiming of debt from Freiherr Georg Wilhelm von Massenbach and the petitioning to Karlsruhe, an actor in this case, regarding an excerpt from the knights’ cadastre of the Canton of Kraichgau are the central activities in the example. Both are implicitly encoded in the <unittitle> along with a couple of other historical facts. The date provided in the <unittitle> is most likely the date on which the petition was created. However, this interpretation is tentative and can only be validated by an archivist or by looking at the actual document.

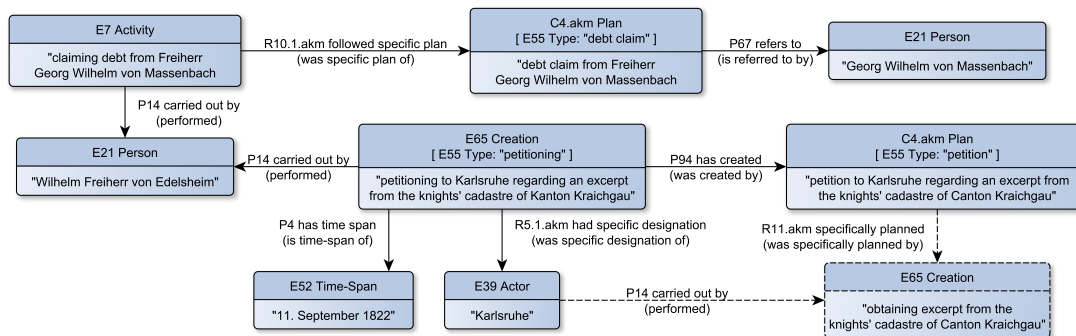


Figure 88 – The information from the <unittitle> represented explicitly (EAD file “labw-f-1-603494”).

The length of the entry in the <unittitle> could be also interpreted as an indication of the lack of suitable options for a more differentiated and explicit representation of historical context information. The semantic overload of entries is also apparent in the case of the three <indexentry> elements, which carry a lot of information.

The entities given with dotted lines in Figure 88 represent possible future occurrences in relation to the events described in the archival pattern: an activity of obtaining the actual excerpt for which the petition is hoping. The assertion is purely hypothetical and is only meant to demonstrate how a history of planning and execution of plans may evolve.

In contrast to the previous example, this case shows a fairly rich description of the historical aboutness context utilizing primarily the *Plans* pattern. The very detailed description is solely based on the information available in the XML snippet and shows the amount of implicit knowledge potentially hidden in metadata fields such as <unittitle>.

Future work could focus on metadata fields such as <unittitle> or <indexentry> and systematically interpret the linguistic and ontological structure of the information these contain, subsequently devising a method for their automatic translation into a data structure.

### Example “Baader Meinhof Trial”

The last example shown in Figure 89 is taken from a finding aid in the LaBW and describes a more recent event: the trial against leading members of the Red Army Faction (*Rote-Armee-Fraktion*, RAF) in the 1970s. The <unittitle> indicates at least two historical facts: the activity of preparing the Baader Meinhof Trial (*Baader-Meinhof-Prozess*) in Stuttgart-Stammheim and its actual conduct. The information provided in the <unitdate> gives the time-span to which the respective files pertain. Without adequate background knowledge of the historical events as well as of the contents of the files described, the exact interpretation of the relationship of the two activities with the files and the indicated time-spans is only preliminary. On the other hand, this example demonstrates once again the importance of explicit representation of knowledge.

```

<c level="file" id="labw-1-86568">
  <did>
    <unitid>EA 1/107 Bü 345</unitid>
    <unittitle>Vorbereitung und Durchführung des Baader-Meinhof-Prozesses in Stuttgart-
      Stammheim</unittitle>
    <unitdate normal="1974-01-01/1975-12-31">1974 - Juli 1975</unitdate>
    <physdesc>
      <extent>1 Bü</extent>
      <dimensions>Folio (Höhe x Breite)</dimensions>
    </physdesc>
    <unitid type="Altsignatur">3012/9</unitid>
  </did>
  <otherfindaid>
    <extref xlink:role="url_archivalunit" xlink:href="http://www.landesarchiv-bw.de/
      plink/?f=1-86568">Titelaufnahme im Angebot des Archivs</extref>
  </otherfindaid>
</c>

```

*Figure 89 – XML-Snippet from EAD file “labw-f-1-5721”*

Figure 90 shows a tentative interpretation of the information provided in the description. The trial is well-known and documented and the archivist would therefore be able to easily identify the activity Baader Meinhof Trial as well as provide additional background information such as other actors involved.

Most importantly, the representation explicitly differentiates between the plan and the act of planning, and the document as the result of an act of documentation, all of which refer to the activity “Baader Meinhof Trial”. The activities of documentation and planning resulting in the archival materials contained in this particular record can be said to have happened during the general time-span “1974 - July 1975” as given in the <unitdate> element. Again, the dates do not necessarily pertain to the trial itself; in fact, the trial was officially begun on 21st May 1975 and the verdict was declared on 28th April 1977, upon which the trial was concluded.

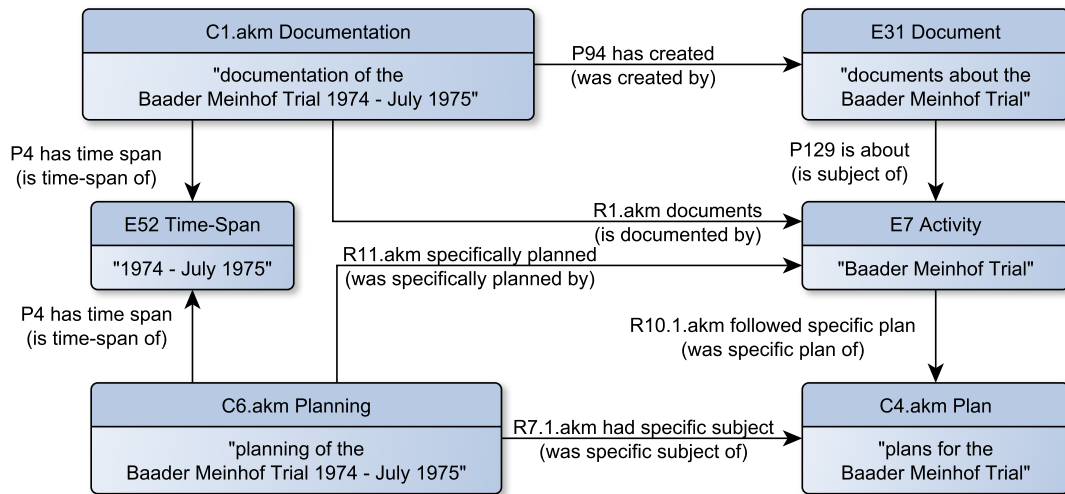


Figure 90 – The information from the <unittitle> represented explicitly (EAD file “labw-f-1-5721”).

While the previous example demonstrated how much implicit information may be contained within a single <unittitle> element, this example shows that in the case of well-known and documented activities and events, additional contextual information can easily be provided by an archivist, at least in terms of intellectual effort.

### 3 Summary: Querying and Implementation

All examples show that a large number of historical facts can be expected to be compressed or implicit in text values of many metadata fields. Further, the examples suggest that archivists are in possession of more knowledge during documentation than is actually encoded, but which could easily be captured by the AKM. In other words, while a full and systematic analysis of EAD would exceed the scope of this study, archival data encoded in EAD as well as the background knowledge of archivists can be expected to provide sufficient information with which to instantiate the AKM at least to a large extent.

The AKM may serve as a basis from which to analyze existing metadata schemas or documentation guidelines and may result in their partial revision in terms of finer-grained semantics or formal structures. For example, based on the *Provenance* pattern, the encoding of the creator, producer and keeper of a document or file could be rendered mandatory. The AKM may also serve as the basis from which new schemas or descriptive layers can be created as additions or supplements to existing and already implemented archival knowledge bases.

The four examples discussed here are, of course, only selective and pertain to individual records. The potential of semantic networks, for which the AKM describes the relevant nodes

and relationships, then lies in the linking of such individual descriptive records within as well as between archival collections, even from different archives. Pattern-directed searches (Dworman et al., 2000) and other complex discovery queries based on the subject matter relevant to archival users become possible.

In general, the universals identified in this study are buried beneath many different words and terms. The AKM allows for the standardization of vocabulary leading to controlled access points, which may be directly derived from the diagrams. If appropriate cataloguing rules existed for the entities described in the AKM, archival materials could be interlinked in a meaningful way and alongside existing archival descriptions.

The identity and identification of particulars; that is, of instances of the classes described in the AKM, is a question of appropriate cataloguing rules which would permit the finding and description of appropriate *controlled access points* which represent “a name, term, code, etc., under which a bibliographic or authority record or reference will be found” (Bourdon et al., 2013, 14) and which “enable a given instance of a given bibliographic entity to be consistently referred to in a given bibliographic database” (Bekiari et al., 2010, 127).

FRBRoo, for example, formally describes the creation of controlled access points (Bekiari et al., 2010, 127-129) which “consists in selecting and assembling existing appellations so as to make the resulting construct as specific, accurate, and ‘unique’ as possible, so as to disambiguate the way a given instance of a given bibliographic entity is consistently referred to in a given bibliographic database” (Bekiari et al., 2010, 127). The person descriptions provided in the third example, for instance, are not only a source for historical knowledge but also for controlled access points which would then allow interlinking with other records. Whether there will be rules for controlled access points for events and what shape these might assume is a question of extended cataloguing rules and documentation guidelines.

Apart from the identification and creation of controlled access points, another important issue is the question of how to query complex semantic networks. Since the AKM is formalized using the CRM which, in turn, has been implemented as an RDFS/OWL representation<sup>97</sup>, the general patterns are immediately implementable as semantic networks. If a particular knowledge structure is given as shown in the diagrams then appropriate data structures can be created from the perspective of information technology.

Tzompanaki and Doerr (2012a) show how large and complex semantic networks may be queried using the CRM. Especially in cases where relevant documents and facts can be expected to be distributed among records or holdings, such patterns would provide relevant access points and contexts from which to retrieve relevant documents or facts. Alexiev (2011) has further discussed how complex semantic networks based on the CRM can be searched by utilizing fundamental relationships and concepts (Tzompanaki and Doerr, 2012b). Zenz et al. (2009) have described a tool called Quick which is an example of a guided method by which users can create progressively semantic queries while still retaining much of the convenience of a keyword

---

<sup>97</sup> <http://erlangen-crm.org/>



search. The tool Nitelight (Russell et al., 2008) is another example of a graphical tool supporting users in query construction and visualization. While only singular examples, these studies show that complex and semantic querying are within the grasp of information systems in the near future. As with all digital information systems that provide search and retrieval facilities, the instruction of users is necessary to a certain degree, at least if the user wishes to comprehend all search options and, most importantly, how results sets are created.

Remaining questions, apart from the previously discussed issues of controlled access points and querying, include how to instantiate a knowledge base that implements the AKM either completely or partially. Archivists could conduct *proactive documentation* of archival collections based on new or revised documentation guidelines and cataloguing rules. As indicated above, if such guidelines and rules indeed existed, only little additional effort would be expected in terms of intellectual work. The primary challenges lie in developing appropriate technical implementation facilitating the entry of data into a knowledge base during documentation.

*Community-driven* approaches in the context of the Web 2.0 (Palmer, 2009; Theimer, 2011) as well as more generally collaborative stances of archives towards external parties (Evans, 2007) have also been discussed by the archival community. Especially regarding general historical expertise surrounding the history of archival materials, crowd-sourcing has a higher potential of providing more additional knowledge than any archivist. The biggest challenges here lie in the social and technical organization of such input as well as the mustering of necessary resources.

*Natural language processing* (NLP) techniques that provide semi-automatic means are likely to facilitate the creation of knowledge bases built on ontologies such as the AKM in the near future. Recent progress in NLP and the many related fields of automatic information detection, extraction and structuring (Clark et al., 2012) permit the confident assumption that semi-automatic or automatic conversion algorithms will be developed for the exploitation of existing archival descriptions and of increasingly available digitized full-text archival resources in order to provide enhanced support for archivists and information seeking users as suggested by the AKM.

There are increasing numbers of projects which have built large and complex semantic networks and knowledge bases for culture-historical research. These projects often utilize the CRM and employ various NLP techniques, especially from the Semantic Web domain. Again, only very few examples from the vast body of research shall suffice to give an impression of the possibilities, such as Nenkova and McKeown (2011) who provided a general overview of automatic summarization techniques or automated annotations as discussed in Ruotsalo et al. (2009). Byrne and Klein (2009) presented automatic (archaeological) event detection and extraction in natural language documents, particularly in the case of the Semantic Web (Byrne, 2009), techniques which would be pivotal to automatically identify and instantiate events and activities in the AKM. The project report on the FDR/Pearl Harbor Project by Ide and Woolner (2004) also demonstrates how historical events can be identified and extracted from historical documents using Semantic Web technologies. Banerjee and Johnson (2015)

recently investigated Linked Data-based entity extraction techniques in order to enhance access to archival collections and found that, even though various quality issues prevail, there is potential for future improvement of, for example, quicker subject analysis and integration of knowledge bases.

A particular example of a culture-historical semantic network is described in Riechert et al. (2010) who have built, similar in purpose to the AKM, an ontology for structuring a prosopographical knowledge base, the *Catalogus Professorum Model* (CPM). The purpose of the project is to create a knowledge base about prosopographical research that can be subjected to network analysis, for instance. The scope of the CPM, therefore, is very different to that of archival descriptions and the scope of the AKM. Furthermore, their knowledge engineering methodology has been based on discussions with historians and the observation of their information behaviour in the semantic editing tool OntoWiki, which was adapted and developed further during the project's lifetime. This is an example of the methodological drawbacks of observational and interrogative means for eliciting user and information needs as discussed in chapter II.

Further examples of large semantic networks have been discussed in Ockeloen et al. (2013), who have built BiographyNet, a semantic knowledge base created "by extracting links between people, historic events, places and time periods from a variety of Dutch biographical dictionaries" (Ockeloen et al., 2013, 1). Another semantic network for historical research has been created by the project "histcross" (Kalus, 2007), which focuses on network analysis and strives to enable qualitative and quantitative historical data analysis and interpretation (Kalus, 2006).

Even though these knowledge bases and semantic networks have different scopes and purposes than those envisaged by the AKM, the previous examples of applied research indicate that the necessary technology is within reach and that knowledge is most likely to be available today in existing archival metadata structures as well as on the part of the archivist. In the short-term, proactive and manual documentation can be implemented for smaller, high-value collections. Such projects could become incubators for the further development of finer-grained metadata schemas, documentation guidelines and cataloguing rules based on information needs as they are ontologically represented in the AKM. Analysis of the AKM could also lead to the short-term revision of existing metadata schemas such as the EAD by identifying elements which could be added or simply made mandatory, such as the creator of archival materials. In any case, the AKM and the empirical work conducted here contribute an indispensable prerequisite for any such endeavour.

## Chapter VII

# Conclusion

### Retrospection

Research on the information behaviour of archive users has repeatedly shown that access to archival knowledge bases such as via finding aids is hindered by a certain deficiency in knowledge about users and their information needs. Archives and their extensive and ever-growing wealth of information and evidential potential constitute a key resource for global knowledge networks (Doerr and Iorizzo, 2008). Facilitating access to archival information systems, as well as improving their search and retrieval tools, demands an extended and deeper empirical understanding and ontological formalization of the behaviour and information needs of archive users.

This study has conducted an analysis of archive user behaviour and requirements, focusing on archival information needs as expressed in reference questions submitted to archives. Such written natural language inquiries sent by mail or email to archives constitute empirical research data containing relatively unfiltered information needs expressed in the users' own words. Interrogative and observational methods of collecting research data on information needs generally elicit either intuition or reactions to a particular system.

The primary research question asked whether it is possible to create an ontology that can reasonably represent archival inquiries and their probable interpretations as formal queries to the model of the archival target world that would generate an adequate response. The hypotheses of this study were that such written natural language inquiries can be interpreted and analyzed in terms of common and shared interests pertaining to archives, that these shared interests can be highly generalized and abstracted to an ontological representation, and that the ontology CIDOC CRM (CRM) would provide adequate means for their formalization.

The primary results are (1) a methodological approach that advocates an interpretative analysis of written natural language inquiries submitted to archives and, as the result of its practical and successful application, (2) the *Archival Knowledge Model* (AKM), an ontological representation of the subject matter which typically appears as the interest of archival inquiries. The successful application of the methodological approach to the analysis of archival reference questions confirms the first two hypotheses and, to a large extent, the third. The CRM has been extended with compatible entities wherever its semantics were found to be inadequate.

### Interpretative Analysis

The interpretative analysis can be formally differentiated into a linguistic analysis and an ontologically oriented phase. The linguistic analysis – understood in the sense of returning to the semantic level of the immediate meaning of words in a phrase – categorizes according to whether archival or non-archival materials, factual information or other services are requested,

and allows for an initial appraisal of the interest of the inquiries. The ontological analysis leaves the immediate level of utterance of the inquiries and further interprets them in order to identify shared interests and to formalize and explicate the subject matter of these interests to common and abstract ontological conceptualizations. Formally, the interpretative analysis can be understood as an act of knowledge engineering, and the process of ontological modelling in particular as an act of ontology engineering.

The analysis is strictly based on the empirical material and the information needs as expressed in the inquiries and is not derived from theory. Beyond the information needs found in the empirical material, the interpretative analysis is mainly guided by two epistemological foundations which are also closely intertwined, the historical and archival domain. Further, the ontology CRM has not only been positively evaluated and instrumental to the ontological modelling but it has also determined to a certain extent the interpretative process; for example, and most importantly, the fact that history is seen as revolving around the notion of events. Finally, the expertise of the author as a historian with extensive background knowledge provided the necessary scholarly foundation for an adequate interpretation of the inquiries.

The methodological approach may also be considered similar to the hermeneutic historical method (IV:2.1) of heuristics, critique and synthesis. The interpretation of the inquiries is a hermeneutic act; only its interest is not derived from original research interests of historical scholarship, and its outcome, the synthesis, is not only reported in prose as a narrative but explicitly formalized in an ontological model.

This study has successfully demonstrated in the practical application of the methodological approach that the subject matter of the interest of inquiries can be brought to a shared and abstract ontological representation. All inquiries can be related to a common ontology, a shared discourse, and highly generalized. The many different inquiries and the phenomena they describe can be reasonably and significantly reduced. Even though the CRM primarily describes a material discourse originating from the museum domain, the CRM has been found adequate for the ontological formalization of the archive user needs interpretatively identified in inquiries. The CRM has been partially extended by six compatible classes and sixteen relationships.

### **Archival Knowledge Model**

The AKM is an ontology which contributes to a deepened empirical understanding of archive users and their information needs. It formally and explicitly represents the subject matter, which typically appears as the primary interest of written natural language inquiries submitted to archives. As such, the AKM describes the historical and archival domain of discourse as represented by the inquiries analyzed and provides a new access perspective to archival holdings which is independent from, yet compatible with, archival descriptions.

As an ontology, the AKM strives to establish a common ontological framework in order to reconcile the conceptualizations of archive users as found in their inquiries with the conceptualizations of the archival domain as provided in archival aids, especially finding aids,

and formalized in schemas and data structures of information systems as well as the implicit knowledge typically available to archivists at the time of documentation. Potential archival aids supplemented by the AKM would provide a more effective structuring of the same knowledge and probably without increased research effort. The archival patterns (VI:2) provide indications (IV:1.2.2) as to the feasibility of such documentation, either based on proactive documentation or on semi-automatic means.

Table 16 summarizes the ten general patterns which constitute the AKM. Each general pattern aggregates a specific constellation of classes and properties describing a prototypical segment of the historical reality to which the subject matter of the interest of inquiries pertains. These aggregations represent ontological representations of user requirements of archival holdings in the form of information needs.

The general patterns can be intellectually distinguished according to their primary epistemic focus, which is either the context of provenance of things (the actor who created, produced or kept things), or the context of aboutness of things (the entities to which archival materials refer). However, query patterns, which exemplify potential queries against an archival knowledge base, would always utilize entities of two general patterns, one from each context.

Table 16 – The general patterns according to their primary epistemic focus.

<b>Provenance Context</b>	<b>Aboutness Context</b>
<i>Provenance (V:2.1.1)</i>	<i>Aboutness (V:2.2.1)</i>
<i>Documentation (V:2.1.2)</i>	<i>Events (V:2.2.2)</i>
<i>Self-Documentation (V:2.1.3)</i>	<i>Plans (V:2.2.3)</i>
<i>Correspondence (V:2.1.4)</i>	<i>Actors (V:2.2.4)</i>
	<i>Things (V:2.2.5)</i>
	<i>Documents (V:2.2.6)</i>

The specific general patterns describe and stress the relevance of various entities; that is, of classes and relationships, such as the explicit documentation of particular persons and particular activities, or that an essential basic interest of users pertains to the provenance of things (who created, produced or kept them, or who might have written something), and that the aboutness appears initially less important.

The purpose of the AKM is not to provide a direct answer to an inquiry but to guide archival users to potentially relevant archival materials which may contain factual or evidential information contributing to finding such an answer. The general patterns provide an additional descriptive access layer which would lead into the archival description, the finding aid, from where the actual identifier or call number of a file or record would be retrieved. Archival principles such as original order or provenance remain relevant as conventional and traditional

access parameters and are neither concerned nor challenged by the AKM.

In other words, the AKM provides a contextual description of a historical context of materials that goes beyond and enriches those typically found in finding aids. The contextual description specifying the path to the material within the archive remains with the finding aids. Direct, factual answers are only of secondary importance to the AKM. The ultimate goal of the AKM is to augment and improve the capabilities of search and retrieval systems by introducing an additional descriptive layer based on the empirically founded analysis and ontological representation of archival information needs.

Simply put, the AKM is essentially concerned with the basic question as to who might have written about a particular historical occurrence of interest, and not with specific information and its location within a specific document. In a similar manner, archival description employs the same necessary simplification by focusing on the principle of *respect des fonds* (IV:2.2). The historian, of course, will further engage with the various documents and create a far more intrinsic and complex knowledge network. Both the AKM and the archival description, then, strive in combination to provide relevant and adequately simplified contextual knowledge structures that will support the historian in obtaining relevant primary sources.

### **Further Outcomes and Discussion**

The will and plan constitute two particular findings of this study and thus two specific insights into archival user needs. The significance of these calls for a brief elaboration on their function here. The *Plans* pattern (V:2.2.3) specifically describes the planning of activities and their plan-based execution, and more specifically the discussion and creation, deciding on and abandoning of plans, as well as the beginning and ending of a will to execute a plan. Furthermore, the plan and the expression of will to act upon and to execute a plan are distinguished from one another.

These distinctions allow us to generally address questions – that is, to identify archival materials that may provide potentially relevant evidential or factual information – about plans and their enactments, and in particular, related to the intentions and aims of actors. Further, they permit insights as to the extent to which plans have been applied, executed and desired (that is, to what extent the intention to execute a plan conforms to historical knowledge or the records regarding the actual execution and enforcement of the same plan). The plan in the form of the more authoritative mandate further plays an important role in the context of documentation and self-documentation activities, each of which are covered by respective general patterns.

The ontological representation of psychological entities such as the will and the more detailed representation of the concepts of planning and plans as well as of more authoritative mandates constitute the primary shortcomings of the CRM in the context of this study. Appropriate and compatible extensions have been suggested. The scope notes of all new classes and properties are provided in the appendix.

The AKM thus permits the characterization of archival materials and their contents according to a clear and fundamental tripartition into the temporal perspectives of the future (the

documented structure of plans), the past (the documented structure of acts), and documented psychological states (the intentions and wills associated with what has been planned and carried out).

The interpretative analysis has also yielded tentative insights into the process of knowledge creation in archival and historical inquiry. The three levels of interest (V:1.3) identified – material fact, psychological and collective statistical – provide indications as to different levels of specificity of the interest in archival and historical inquiries and different stages in the knowledge creation process. Furthermore, archive users appear able to break down information requests into material facts, which indicates that researchers can operate on such a factual level and that the method of factual documentation as suggested by the AKM is indeed both appropriate and effective.

This study further indicates that core metadata exist, understood, however, not as a plain set of a few single entities such as the Dublin Core<sup>98</sup> but as rich semantic patterns of classes and relationships. Yet the subject matter which appears as the typical interest of inquiries can be generalized and abstracted to a relatively small set of classes and relationships, and further aggregated into general patterns. Furthermore, since the inquiries analyzed data from a time-span of about 15 years, there are indications that the interests of users, in other words the identifiable shared subject matter, or the real questions, appear not to change significantly over this kind of time period.

The outcomes of this analysis increase empirical confidence regarding the nature of questions submitted to archives and the information needs of archive users, since the general patterns are firmly rooted in empirical evidence in the form of original written natural language inquiries. The AKM may then serve as an intellectual guide for the evaluation and migration of existing — and the design of future — data structures as well as documentation guidelines. For example, the AKM may be used relatively directly for the evaluation of archival metadata standards such as the Encoding Archival Description (EAD) by determining the relevance of particular metadata fields and data values, as well as their potential for improving search and retrieval. As a rule, the general patterns provide empirically founded indications of possible improvements such as relevant semantic links between archival items and particular information entities, and meaningful trails through the body of archival materials and records. The initiation and actual implementation of such improvements, however, remains the task of future projects and committees, beginning with the assessment of prospective costs, necessary resources and further operationalization of the AKM. While these aspects exceed the scope of this analysis, with the AKM, however, the study at hand has successfully established a constitutive logical and theoretical framework for future work in this field.

Supplementing archival descriptions with additional access points to archival materials which embody conceptualizations of users will increase the value and attractiveness of the holdings. Users would more frequently and readily conduct their own searches, and “accidental

---

<sup>98</sup> <http://dublincore.org/>

users” (Craig, 1998, 125) may be drawn in more easily and effectively. As the archival patterns suggest, in terms of the necessary knowledge that needs to be explicitly formalized and encoded in order to meet the requirements of typical and relevant user interests, the archivists is most likely to obtain and possess such knowledge during documentation. Adding such information to archival finding aids would therefore require only a limited and feasible additional effort.

### **Future Work and Outlook**

Both the AKM and the methodological approach presuppose further research into various specific aspects of their disposition as well as providing the foundation for related follow-up work.

One of the next steps regarding the evaluation and standardization of the AKM is the submission of the extensions made to the CRM to the CRM SIG committee (IV:3.1). These extensions would enter an official process of deliberation by experts, possibly undergo revision, and then attain formal recognition.

Despite the high value of original written natural language inquiries, one of the major obstacles has been to obtain access to and collect reference questions. Access is highly dependent on national data privacy laws and on the local organization of reference services in the archive. Internal procedures at the archive before access is granted can take a long time. As these obstacles are difficult to circumvent, researchers need to be aware of them and address such issues proactively. Freely accessible databases such as the OEP (III:2.2) may, however, greatly facilitate the collection and selection process.

Furthermore, the intellectual and manual conduct of the iterative interpretative analysis was both cumbersome and time-consuming. Since the initial development of the interpretative approach has itself been an objective of this study, the adaptive and applied processing of the collected reference questions has been both essential and instrumental. However, with the methodological approach established and substantial confidence regarding its validity and adequacy, further investigation into feasible semi-automatic or even automatic means that would facilitate future research and follow-up studies employing, and possibly further developing, the same methodological approach would appear fruitful. For example, natural language processing techniques should be considered in order to semi-automatically or even automatically analyze user inquiries on a larger scale.

Furthermore, additional samples need to be collected and analyzed in order to evaluate and further validate the methodological approach as well as the specific outcomes. Samples may, but need not, cover a similar topical and temporal horizon to those analyzed here. For example, one sample may focus on a more specific topic or historical period; another may collect inquiries directed to a particular archive department. The period during which the inquiries have been posed to an archive may also constitute valuable selection criteria with which to investigate whether general patterns, and thus fundamental interests of the archive users, significantly change over time. The feasibility of each option, however, depends on the extent and scope of



the internal documentation regime and registry system employed at the respective archive.

Additional user studies may possibly result in the extension, or even partial unification of entities defined by the AKM. By conducting more studies, the AKM and its general patterns may be extended or limited in terms of adding or removing, or by generalizing or specializing some of their classes and relationships. For example, the differentiation between conceptual and physical items is certainly relevant and probably even mandatory from a logical and theoretical point of view; however, only the actual implementation and application of the AKM and additional user studies could confirm its practical relevance.

Additional general patterns may emerge as specializations or combinations of those existing. For example, even though the designation of documents to a “public” or “internal” audience can be represented using the *Correspondence* pattern (V:2.1.4), its relationship to the *Self-Documentation* pattern (V:2.1.3) could be further investigated. In this context, the nature of “official” and “unofficial” documents also deserves additional attention. Future research would further need to determine potential simplifications and unifications of single classes and relationships. Furthermore, their formal and machine-readable representation constitutes another important step in the development of the AKM. For this purpose, existing RDFS/OWL implementations could be used and extended. As such, the AKM has established the initial logical and theoretical framework.

While the shared subject matter of the interests of the inquiries can be highly generalized and reduced to a relatively small set of classes and relationships, the instantiation of these, as indicated by the query patterns, allows us, in turn, to address a virtually unlimited number of individual questions and inquiries. Here, research could explore the instance level and the nature of particulars identified in specific cases and application scenarios. Furthermore, research may study the terminological level of particular types and try to identify vocabularies of controlled terms or categories. Both controlled vocabularies of types as well as of class instances will be important for the practical realization of archival information systems based on the AKM. The scope of this study as well as of the AKM pertains to the ontological level. For this reason, no attempt has been made to further distinguish the ontological entities of the AKM into particular types or to specify particular instances of classes.

In general, continued research is required on archival reference questions, which constitute an information-rich source of primary research data for analyzing the information behaviour of archive users from various perspectives. In particular, more extensive and dedicated investigation is to be encouraged into the aforementioned levels of interest during the archival and historical knowledge creation process, as well as the relationship between the questions of users and the answers of the archivists. Archives hold large quantities of such research data that would allow continuous and systematic as well as empirical investigation into the nature of questions and the information needs of their users.

Archive reference questions as empirical research data should also be rendered more easily accessible by archives for external research. Depending on the extent of the records, user files

would, for example, allow the investigation of the evolution of information needs during the user's engagement with the archive, or shed light on internal procedures and communications during the processing of user inquiries by archivists. Of course, data protection and privacy issues need to be taken into account.

The methodological approach to the interpretative analysis of archival inquiries in this study inspires future qualitative research that approaches users and their information needs more thoroughly and from a different perspective than currently found in archival research, thus focusing more on adequate data structures.

In terms of the research data used and the methodological approach applied, the research presented in this study appears to be rare among work on the information behaviour of archive users. Archival reference questions are research data that is difficult to obtain and analyze; however, the interpretative analysis and formalization of written natural language questions from users to archives, as has been demonstrated here, constitute a valuable source of meaningful data on original users and their information needs. It is only by gaining a deeper understanding of, as well as a consensus on, archive users and their information needs that new generations of more sophisticated pattern-oriented information systems will be able to serve their users with their real questions.

# Appendix

## Definition of the AKM

The following four tables list all classes and properties that constitute the Archival Knowledge Model (AKM). The scope notes of CRM classes and properties are taken from the “Definition of the CIDOC Conceptual Reference Model”, Version 5.0.1, March 2009 (Crofts et al., 2006) without the additional notes on constraints, sub-classes and sub-properties, and examples. One class and one property have been taken from FRBRoo, Version 1.0.1 (Bekiari et al., 2010).

### Native AKM Classes

The table lists all new classes been created for the AKM.

Table 17 – Native AKM Classes

<b>ID</b>	<b>Class Name</b>	<b>Scope Note</b>
C1 <sub>akm</sub>	<i>Documentation</i>	<p><i>Subclass of E65 Creation</i></p> <p>The class comprises instances of activities (E7) which observe or document, in the most general sense, events (E5) or, in most cases, activities (E7) conducted by actors other than those performing the documentation activity.</p> <p>The documentation or observation may be direct or indirect, and a-posteriori in relation to the time and place of the documented event or activity.</p> <p>The documentation or observation may subsume a series of direct and indirect acts of observation and documentation spread out over a period of time but which appear overall as one coherent activity. The definition of what forms a coherent documentation activity, its extent and cohesion in terms of geography, time, participation and objective, is a question of reasonable abstraction.</p> <p>Documentation activities are carried out by one or more individual persons (E39) or groups (E74), and typically result in one or more documents (E31) containing an account of the documented events or activities of others.</p> <p>A mandate (C5<sub>akm</sub>) may determine the principle objective or target, a particular actor or type of actor, of a documentation activity, as well as the actor(s) formally tasked with its execution. This actor does not necessary need to be the same actor who actually conducted the documentation activity.</p> <p><i>Examples</i></p> <ul style="list-style-type: none"><li>• the documentation of the birth of a human being by issuing a birth certificate</li><li>• the creation of a dossier about the past activities of a group of people</li><li>• the act of surveillance of the activities of a person</li></ul>

ID	Class Name	Scope Note
C2 <sub>akm</sub>	<i>Self-Documentation</i>	<p data-bbox="571 322 890 349"><i>Subclass of C1<sub>akm</sub> Documentation</i></p> <p data-bbox="571 353 1418 412">The class comprises instances of activities (E7) which observe or document activities (E7) conducted by the same actor who also conducts the self-documentation activity.</p> <p data-bbox="571 416 1418 537">The main difference to documentation activities (C1<sub>akm</sub>) is the direct or indirect observation and documentation of one's own activities. The results of (mandate-based) self-documentation activities are typically official accounts (<i>Rechenschaftsberichte</i>) such as proceedings, government statements or minutes.</p> <p data-bbox="571 542 1418 600">All other aspects of the scope note of documentation (C1<sub>akm</sub>) also pertain to self-documentation (C2<sub>akm</sub>).</p> <p data-bbox="571 604 655 631"><i>Examples</i></p> <ul data-bbox="616 636 1418 734" style="list-style-type: none"> <li data-bbox="616 636 1418 694">• the members of a parliamentary committee document their meetings through minutes</li> <li data-bbox="616 698 1418 734">• the participants of a journey create a report about their trip</li> </ul>
C3 <sub>akm</sub>	<i>Expression of Will</i>	<p data-bbox="571 770 895 797"><i>Subclass of E73 Information Object</i></p> <p data-bbox="571 801 1418 1016">The class comprises instances of information objects (E73) which contain the symbolic form of the temporal entity (E2) will. The will is a psychological entity which constitutes the inner conviction of an actor to act. As such, a will is volatile and not directly observable or recordable. When a will is expressed or externalized by an actor in any form, then a will in a specific state can be recorded in symbolic form as an expression of will. Wills may be expressed and externalized by an actor in different forms foremost by speech or writing, but also by graphics, gestures, or behaviour.</p> <p data-bbox="571 1021 1418 1111">A will may be expressed by individual persons or by groups. A parliamentary body, for example, may express its will to enact a law by passing it by vote or a newly elected chancellor may express a will to change existing laws during an inaugural speech.</p> <p data-bbox="571 1115 1418 1267">Laws and legislative proposals are instructive examples as to why the plan and the expression of will are distinguished. While parliamentary committees prepare the legislative proposal, only the parliamentary body performs an act of will by either passing or by rejecting the proposal. This act of parliament (C3<sub>akm</sub>) is legally separate from the preparatory deliberation (C6<sub>akm</sub>) of the legislative proposal (C4<sub>akm</sub>).</p> <p data-bbox="571 1272 655 1299"><i>Examples</i></p> <ul data-bbox="616 1303 1046 1370" style="list-style-type: none"> <li data-bbox="616 1303 1046 1330">• a law taking effect</li> <li data-bbox="616 1335 1046 1370">• the American Declaration of Independence</li> </ul>
C4 <sub>akm</sub>	<i>Plan</i>	<p data-bbox="571 1406 906 1433"><i>Subclass of E29 Design or Procedure</i></p> <p data-bbox="571 1438 1418 1653">The class comprises instances of information objects (E73) which contain the symbolic form of the intellectual structure (E89) of a plan. This intellectual structure, or "plan structure" (<i>Planstruktur</i>), constitutes a schema for the execution of activities (E7). Three general types of plan structures can be distinguished: (a) <i>descriptions of appearance</i>, for example a plan of a building or a topographical map, (b) <i>descriptions of procedure</i>, for example a legal regulation, law or mandate, and (c) <i>descriptions for achieving a condition</i> or future state of affairs, for example a political agenda or policy.</p> <p data-bbox="571 1657 655 1684"><i>Examples</i></p> <ul data-bbox="616 1688 1254 1785" style="list-style-type: none"> <li data-bbox="616 1688 1254 1715">• the social policy of the British government in the 19th century</li> <li data-bbox="616 1720 1254 1747">• the program of the SED to solve the housing problem in the GDR</li> <li data-bbox="616 1751 1254 1785">• the building plan of the ground floor of the Reichstag</li> </ul>

ID	Class Name	Scope Note
C5 <sub>akm</sub>	<i>Mandate</i>	<p data-bbox="571 315 788 347"><i>Subclass of C4<sub>akm</sub> Plan</i></p> <p data-bbox="571 349 1415 443">The class comprises instances of information objects (E73) which contain the symbolic form of the intellectual structure (E89) of a plan that describes the principal scope of documentation activities (C1<sub>akm</sub>).</p> <p data-bbox="571 445 1415 535">In contrast to instances of plans (C4<sub>akm</sub>), mandates are explicitly authoritative and trans- pose an intention as well as the authorization to conduct an act, typically in a regular and routinely manner.</p> <p data-bbox="571 537 1415 723">Documentation activities are typically based on or follow mandates such as formal codes of responsibility, by-laws, internal rules and procedures, or standing orders. Such mandates may specify (a) the targets of a documentation activity such as a particular or type of actor, event or activity, and (b) the actor who is formally tasked or entrusted with carrying out such activities. Both aspects set the mandate apart from the plan (C4<sub>akm</sub>). A mandate may also be given ad hoc, such as in the case of an order on the battle field.</p> <p data-bbox="571 725 659 757"><i>Examples</i></p> <ul data-bbox="616 768 1415 920" style="list-style-type: none"> <li data-bbox="616 768 1415 826">• The police conducts surveillance on a specific group based on a court order which legalizes the surveillance and defines what is allowed and what not;</li> <li data-bbox="616 828 954 860">• parliamentary rules of procedure</li> <li data-bbox="616 862 959 893">• internal regulations of companies</li> <li data-bbox="616 896 788 920">• standing orders</li> </ul>
C6 <sub>akm</sub>	<i>Planning</i>	<p data-bbox="571 954 783 985"><i>Subclass of E7 Activity</i></p> <p data-bbox="571 987 1415 1081">The class comprises instances of activities (E7) which (a) have planned or prepared another activity known or assumed to have taken place, or which (b) have discussed or talked about such plans (C4<sub>akm</sub>).</p> <p data-bbox="571 1084 1415 1234">The planning activity may subsume a series of direct and indirect acts of preparation or deliberation spread out over a period of time but which appear overall as one coherent activity. The definition of what forms a coherent planning activity, its extent and cohesion in terms of geography, time, participation and objective, is a question of reasonable abstraction.</p> <p data-bbox="571 1236 659 1267"><i>Examples</i></p> <ul data-bbox="616 1279 1302 1368" style="list-style-type: none"> <li data-bbox="616 1279 1302 1310">• the preparatory deliberation of the Reichstag for a legislative proposal</li> <li data-bbox="616 1312 1158 1344">• the planning meetings of the FDGB for a trip to France</li> <li data-bbox="616 1346 1118 1368">• the German preparation of "Operation Barbarossa"</li> </ul>
C7 <sub>akm</sub>	<i>Finding Aid</i>	<p data-bbox="571 1402 815 1433"><i>Subclass of E31 Document</i></p> <p data-bbox="571 1435 1415 1494">The class comprises instances of documents (E31) which specifically describe and docu- ment any kind of archival holding (E24).</p> <p data-bbox="571 1496 659 1527"><i>Examples</i></p> <ul data-bbox="616 1538 1310 1597" style="list-style-type: none"> <li data-bbox="616 1538 1310 1570">• the finding aid labw-f-1-5721 of the Landesarchiv Baden-Württemberg</li> <li data-bbox="616 1572 1114 1597">• the finding aid "Red Suitcase" of the Bundesarchiv</li> </ul>

## Native CRM Classes

The table lists all classes taken from the CRM (Crofts et al., 2006), whereby one class has been taken from FRBRoo (Bekiari et al., 2010), and reused in the AKM.

Table 18 – Native CRM Classes

ID	Class Name	Scope Note
E1	<i>CRM Entity</i>	<p>This class comprises all things in the universe of discourse of the CIDOC Conceptual Reference Model. It is an abstract concept providing for three general properties:</p> <ol style="list-style-type: none"> <li>1. Identification by name or appellation, and in particular by a preferred identifier</li> <li>2. Classification by type, allowing further refinement of the specific subclass an instance belongs to</li> <li>3. Attachment of free text for the expression of anything not captured by formal properties.</li> </ol> <p>With the exception of E59 Primitive Value, all other classes within the CRM are directly or indirectly specialisations of E1 CRM Entity.</p>
E4	<i>Period</i>	<p>This class comprises sets of coherent phenomena or cultural manifestations bounded in time and space. It is the social or physical coherence of these phenomena that identify an E4 Period and not the associated spatio-temporal bounds. These bounds are a mere approximation of the actual process of growth, spread and retreat. Consequently, different periods can overlap and coexist in time and space, such as when a nomadic culture exists in the same area as a sedentary culture.</p> <p>Typically this class is used to describe prehistoric or historic periods such as the "Neolithic Period", the "Ming Dynasty" or the "McCarthy Era". There are however no assumptions about the scale of the associated phenomena. In particular all events are seen as synthetic processes consisting of coherent phenomena. Therefore E4 Period is a superclass of E5 Event. For example, a modern clinical E67 Birth can be seen as both an atomic E5 Event and as an E4 Period that consists of multiple activities performed by multiple instances of E39 Actor.</p> <p>There are two different conceptualisations of 'artistic style', defined either by physical features or by historical context. For example, "Impressionism" can be viewed as a period lasting from approximately 1870 to 1905 during which paintings with particular characteristics were produced by a group of artists that included (among others) Monet, Renoir, Pissarro, Sisley and Degas. Alternatively, it can be regarded as a style applicable to all paintings sharing the characteristics of the works produced by the Impressionist painters, regardless of historical context. The first interpretation is an E4 Period, and the second defines morphological object types that fall under E55 Type.</p> <p>Another specific case of an E4 Period is the set of activities and phenomena associated with a settlement, such as the populated period of Nineveh.</p>
E5	<i>Event</i>	<p>This class comprises changes of states in cultural, social or physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological or legal phenomena. Such changes of state will affect instances of E77 Persistent Item or its subclasses. The distinction between an E5 Event and an E4 Period is partly a question of the scale of observation. Viewed at a coarse level of detail, an E5 Event is an 'instantaneous' change of state. At a fine level, the E5 Event can be analysed into its component phenomena within a space and time frame, and as such can be seen as an E4 Period. The reverse is not necessarily the case: not all instances of E4 Period give rise to a noteworthy change of state.</p>

<b>ID</b>	<b>Class Name</b>	<b>Scope Note</b>
E7	<i>Activity</i>	This class comprises actions intentionally carried out by instances of E39 Actor that result in changes of state in the cultural, social, or physical systems documented. This notion includes complex, composite and long-lasting actions such as the building of a settlement or a war, as well as simple, short-lived actions such as the opening of a door.
E9	<i>Move</i>	This class comprises changes of the physical location of the instances of E19 Physical Object. Note, that the class E9 Move inherits the property P7 took place at (witnessed) E53 Place. This property should be used to describe the trajectory or a larger area within which a move takes place, whereas the properties P26 moved to (was destination of), P27 moved from (was origin of) describe the start and end points only. Moves may also be documented to consist of other moves (via P9 consists of (forms part of)), in order to describe intermediate stages on a trajectory. In that case, start and end points of the partial moves should match appropriately between each other and with the overall event.
E10	<i>Transfer of Custody</i>	This class comprises transfers of physical custody of objects between instances of E39 Actor. The recording of the donor and/or recipient is optional. It is possible that in an instance of E10 Transfer of Custody there is either no donor or no recipient. Depending on the circumstances it may describe: 1. the beginning of custody 2. the end of custody 3. the transfer of custody 4. the receipt of custody from an unknown source 5. the declared loss of an object. The distinction between the legal responsibility for custody and the actual physical possession of the object should be expressed using the property P2 has type (is type of). A specific case of transfer of custody is theft. The interpretation of the museum notion of "accession" differs between institutions. The CRM therefore models legal ownership and physical custody separately. Institutions will then model their specific notions of accession and deaccession as combinations of these.
E12	<i>Production</i>	This class comprises activities that are designed to, and succeed in, creating one or more new items. It specializes the notion of modification into production. The decision as to whether or not an object is regarded as new is context sensitive. Normally, items are considered "new" if there is no obvious overall similarity between them and the consumed items and material used in their production. In other cases, an item is considered "new" because it becomes relevant to documentation by a modification. For example, the scribbling of a name on a potsherd may make it a voting token. The original potsherd may not be worth documenting, in contrast to the inscribed one. This entity can be collective: the printing of a thousand books, for example, would normally be considered a single event. An event should also be documented using E81 Transformation if it results in the destruction of one or more objects and the simultaneous production of others using parts or material from the originals. In this case, the new items have separate identities and matter is preserved, but identity is not.
E21	<i>Person</i>	This class comprises real persons who live or are assumed to have lived. Legendary figures that may have existed, such as Ulysses and King Arthur, fall into this class if the documentation refers to them as historical figures. In cases where doubt exists as to whether several persons are in fact identical, multiple instances can be created and linked to indicate their relationship. The CRM does not propose a specific form to support reasoning about possible identity.
E24	<i>Physical Man-Made Thing</i>	This class comprises all persistent physical items that are purposely created by human activity. This class comprises man-made objects, such as a swords, and man-made features, such as rock art. No assumptions are made as to the extent of modification required to justify regarding an object as man-made. For example, a "cup and ring" carving on bedrock is regarded as instance of E24 Physical Man-Made Thing.

<b>ID</b>	<b>Class Name</b>	<b>Scope Note</b>
E31	<i>Document</i>	This class comprises identifiable immaterial items that make propositions about reality. These propositions may be expressed in text, graphics, images, audiograms, videograms or by other similar means. Documentation databases are regarded as a special case of E31 Document. This class should not be confused with the term "document" in Information Technology, which is compatible with E73 Information Object.
E35	<i>Title</i>	This class comprises the names assigned to works, such as texts, artworks or pieces of music. Titles are proper noun phrases or verbal phrases, and should not be confused with generic object names such as "chair", "painting" or "book" (the latter are common nouns and are modelled in the CRM as instances of E55 Type). Titles may be assigned by the creator of the work itself, or by a social group. This class also comprises the translations of titles that are used as surrogates for the original titles in different social contexts.
E39	<i>Actor</i>	This class comprises people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible. The CRM does not attempt to model the inadvertent actions of such actors. Individual people should be documented as instances of E21 Person, whereas groups should be documented as instances of either E74 Group or its subclass E40 Legal Body.
E41	<i>Appellation</i>	This class comprises all sequences of signs of any nature, either meaningful or not, that are used or can be used to refer to and identify a specific instance of some class within a certain context. Instances of E41 Appellation do not identify things by their meaning, even if they happen to have one, but by convention, tradition, or agreement. Instances of E41 Appellation are cultural constructs; as such, they have a context, a history, and a use in time and space by some group of users. A given instance of E41 Appellation can have alternative forms, i.e., other instances of E41 Appellation that are always regarded as equivalent independent from the thing it denotes. Specific subclasses of E41 Appellation should be used when instances of E41 Appellation of a characteristic form are used for particular objects. Instances of E49 Time Appellation, for example, which take the form of instances of E50 Date, can be easily recognised.
E42	<i>Identifier</i>	This class comprises strings or codes assigned to instances of E1 CRM Entity in order to identify them uniquely and permanently within the context of one or more organisations. Such codes are often known as inventory numbers, registration codes, etc. and are typically composed of alphanumeric sequences. The class E42 Identifier is not normally used for machine-generated identifiers used for automated processing unless these are also used by human agents.
E52	<i>Time-Span</i>	This class comprises abstract temporal extents, in the sense of Galilean physics, having a beginning, an end and a duration. Time Span has no other semantic connotations. Time-Spans are used to define the temporal extent of instances of E4 Period, E5 Event and any other phenomena valid for a certain time. An E52 Time-Span may be identified by one or more instances of E49 Time Appellation. Since our knowledge of history is imperfect, instances of E52 Time-Span can best be considered as approximations of the actual Time-Spans of temporal entities. The properties of E52 Time-Span are intended to allow these approximations to be expressed precisely. An extreme case of approximation, might, for example, define an E52 Time-Span having unknown beginning, end and duration. Used as a common E52 Time-Span for two events, it would nevertheless define them as being simultaneous, even if nothing else was known. Automatic processing and querying of instances of E52 Time-Span is facilitated if data can be parsed into an E61 Time Primitive.



ID	Class Name	Scope Note
E53	<i>Place</i>	<p>This class comprises extents in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter. The instances of E53 Place are usually determined by reference to the position of "immobile" objects such as buildings, cities, mountains, rivers, or dedicated geodetic marks. A Place can be determined by combining a frame of reference and a location with respect to this frame. It may be identified by one or more instances of E44 Place Appellation. It is sometimes argued that instances of E53 Place are best identified by global coordinates or absolute reference systems. However, relative references are often more relevant in the context of cultural documentation and tend to be more precise. In particular, we are often interested in position in relation to large, mobile objects, such as ships. For example, the Place at which Nelson died is known with reference to a large mobile object – H.M.S Victory. A resolution of this Place in terms of absolute coordinates would require knowledge of the movements of the vessel and the precise time of death, either of which may be revised, and the result would lack historical and cultural relevance. Any object can serve as a frame of reference for E53 Place determination. The model foresees the notion of a "section" of an E19 Physical Object as a valid E53 Place determination.</p>
E54	<i>Dimension</i>	<p>This class comprises quantifiable properties that can be measured by some calibrated means and can be approximated by values, i.e. points or regions in a mathematical or conceptual space, such as natural or real numbers, RGB values etc. An instance of E54 Dimension represents the true quantity, independent from its numerical approximation, e.g. in inches or in cm. The properties of the class E54 Dimension allow for expressing the numerical approximation of the values of an instance of E54 Dimension. If the true values belong to a non-discrete space, such as spatial distances, it is recommended to record them as approximations by intervals or regions of indeterminacy enclosing the assumed true values. For instance, a length of 5 cm may be recorded as 4.5-5.5 cm, according to the precision of the respective observation. Note, that interoperability of values described in different units depends critically on the representation as value regions. Numerical approximations in archaic instances of E58 Measurement Unit used in historical records should be preserved. Equivalents corresponding to current knowledge should be recorded as additional instances of E54 Dimension as appropriate.</p>
E65	<i>Creation</i>	<p>This class comprises events that result in the creation of conceptual items or immaterial products, such as legends, poems, texts, music, images, movies, laws, types etc.</p>
E73	<i>Information Object</i>	<p>This class comprises identifiable immaterial items, such as a poems, jokes, data sets, images, texts, multimedia objects, procedural prescriptions, computer program code, algorithm or mathematical formulae, that have an objectively recognizable structure and are documented as single units. An E73 Information Object does not depend on a specific physical carrier, which can include human memory, and it can exist on one or more carriers simultaneously. Instances of E73 Information Object of a linguistic nature should be declared as instances of the E33 Linguistic Object subclass. Instances of E73 Information Object of a documentary nature should be declared as instances of the E31 Document subclass. Conceptual items such as types and classes are not instances of E73 Information Object, nor are ideas without a reproducible expression.</p>

ID	Class Name	Scope Note
E74	<i>Group</i>	<p>This class comprises any gatherings or organizations of two or more people that act collectively or in a similar way due to any form of unifying relationship. In the wider sense this class also comprises official positions which used to be regarded in certain contexts as one actor, independent of the current holder of the office, such as the president of a country. A gathering of people becomes an E74 Group when it exhibits organizational characteristics usually typified by a set of ideas or beliefs held in common, or actions performed together. These might be communication, creating some common artifact, a common purpose such as study, worship, business, sports, etc. Nationality can be modeled as membership in an E74 Group (cf. HumanML markup). Married couples and other concepts of family are regarded as particular examples of E74 Group.</p>
E78	<i>Collection</i>	<p>This class comprises aggregations of instances of E18 Physical Thing that are assembled and maintained ("curated" and "preserved," in museological terminology) by one or more instances of E39 Actor over time for a specific purpose and audience, and according to a particular collection development plan. Items may be added or removed from an E78 Collection in pursuit of this plan. This class should not be confused with the E39 Actor maintaining the E78 Collection often referred to with the name of the E78 Collection (e.g. "The Wallace Collection decided..."). Collective objects in the general sense, like a tomb full of gifts, a folder with stamps or a set of chessmen, should be documented as instances of E19 Physical Object, and not as instances of E78 Collection. This is because they form wholes either because they are physically bound together or because they are kept together for their functionality.</p>
F33	<i>Reproduction Event</i>	<p>This class comprises activities that consist in making copies, more or less mechanically, of an instance of E84 Information Carrier (such as an F5 Item or an F4 Manifestation Singleton which is also instance of E84 Information Carrier), preserving the expression carried by it. A Reproduction Event results in new instances of E84 Information Carrier coming into existence. In general, the copy will have different attributes from the original and they are therefore not regarded as siblings. This class makes it possible to account for the legal distinction between private copying for the purpose of "fair use," and mass production for the purpose of dissemination. It can prove difficult to determine where to draw the line between F33 Reproduction Event and F32 Carrier Production Event in cases where multiple copies are produced. In this case, the copies, but not the original, may be regarded as instances of F5 Item. It is the existence of an explicit production plan that makes the difference. As a consequence, F33 Reproduction Event<sup>54</sup> and F32 Carrier Production Event are not declared as disjoint, which makes it possible to account for such situations that could be regarded as instances of both Production Event and Reproduction Event.</p>

## Native AKM Properties

The table lists all new properties created for the AKM.

Table 19 – Native AKM Properties

ID	Property Name	Domain	Range	Scope Note
R1 <sub>akm</sub>	<i>documents (is documented by)</i>	C1 <sub>akm</sub> Documentation	E1 CRM Entity	This property describes the relationship between a documenting activity (C1 <sub>akm</sub> ) and another entity (E1) which is being documented, typically an event (E5) or activity (E7), or alternatively an actor (E39).
R2 <sub>akm</sub>	<i>is based on (informs)</i>	E7 Activity	C2 <sub>akm</sub> Mandate	This property describes the relationship between a activity (E7) and the mandate (C2 <sub>akm</sub> ) which governs its conduct in terms of the designated executor (E39) and principle target. A mandate may inform any activity; however, some of the most common cases are documentation activities (C1 <sub>akm</sub> ).
R3 <sub>akm</sub>	<i>was entrusted with (was entrusted to)</i>	E39 Actor	C2 <sub>akm</sub> Mandate	This property describes the relationship between an actor (E39) and the mandate (C2 <sub>akm</sub> ) which has been formally entrusted to said actor.
R4.1 <sub>akm</sub>	<i>has specific target (is specific target of)</i>	C2 <sub>akm</sub> Mandate	E39 Actor	This property describes the relationship between a mandate (C2 <sub>akm</sub> ) and an actor (E39) who is the specific designated target of said mandate.
R4.2 <sub>akm</sub>	<i>has general target (is general target of)</i>	C2 <sub>akm</sub> Mandate	E55 Type	This property describes the relationship between a mandate (C2 <sub>akm</sub> ) and a type (E55) of actor (E39) who is the principal designated target of said mandate.
R5.1 <sub>akm</sub>	<i>had specific designation (was specific designation of)</i>	E65 Creation	E39 Actor	This property describes the relationship between a creation activity (E65) such as writing a letter or issuing a communiqué and the specific actor (E39) who is the designated recipient of said creation activity. The information object (E73) which being created is intended to address a person or group. Whether or not said information object actually reaches its goal is immaterial.
R5.2 <sub>akm</sub>	<i>had general designation (was general designation of)</i>	E65 Creation	E55 Type	This property describes the relationship between a creation activity (E65) such as writing a letter or issuing a communiqué and a type (E55) of actor (E39) who is the designated recipient of said creation activity. The information object (E73) being created is intended to address a person or group. Whether or not said information object actually reaches its goal is immaterial.
R6.1 <sub>akm</sub>	<i>has specific addressee (is specific addressee of)</i>	E73 Information Object	E39 Actor	This property describes the relationship between an information object (E73) and the specific actor (E39) to whom said information object has been addressed. The property expresses a phenomenological relationship such as the address line on a letter or the salutation in a speech.

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
R6.2 <sub>akm</sub>	<i>has general addressee (is general addressee of)</i>	E73 Information Object	E55 Type	This property describes the relationship between an information object (E73) and the type (E55) of actor (E39) to whom said information object has been addressed. The property expresses a phenomenological relationship such as the address line on a letter or the salutation in a speech.
R7.1 <sub>akm</sub>	<i>had specific subject (was specific subject of)</i>	C6 <sub>akm</sub> Planning	C4 <sub>akm</sub> Plan	This property describes the relationship between a planning activity (C6 <sub>akm</sub> ) and the specific plan (C4 <sub>akm</sub> ) which was the subject during said planning activity.
R7.2 <sub>akm</sub>	<i>was generally about (was general subject of)</i>	C6 <sub>akm</sub> Planning	E55 Type	This property describes the relationship between a planning activity (C6 <sub>akm</sub> ) and the type (E55) of plan (C4 <sub>akm</sub> ) which was the subject during said planning activity.
R8 <sub>akm</sub>	<i>decided (was decided in)</i>	C6 <sub>akm</sub> Planning	C4 <sub>akm</sub> Plan	<i>Sub-property of R7.1<sub>akm</sub></i> This property describes the relationship between a planning activity (C6 <sub>akm</sub> ) and the specific plan (C4 <sub>akm</sub> ) decided upon to be enacted during said planning activity; for example, the passing of a law in parliament. Note that in this case the planning activity (C6 <sub>akm</sub> ) is also an instance of C3 <sub>akm</sub> Expression of Will.
R9 <sub>akm</sub>	<i>abandoned (was abandoned by)</i>	C6 <sub>akm</sub> Planning	C4 <sub>akm</sub> Plan	<i>Sub-property of R7.1<sub>akm</sub></i> This property describes the relationship between a planning activity (C6 <sub>akm</sub> ) and the specific plan (C4 <sub>akm</sub> ) decided upon to be abandoned during said planning activity; for example, the rejection of a legislative proposal in parliament. Note that in this case the planning activity (C6 <sub>akm</sub> ) is also an instance of C3 <sub>akm</sub> Expression of Will.
R10.1 <sub>akm</sub>	<i>followed specific plan (was specific plan of)</i>	E7 Activity	C4 <sub>akm</sub> Plan	<i>Sub-property of P33</i> This property describes the relationship between an activity (E7) and a specific plan (C4 <sub>akm</sub> ) upon which said activity is based. The activity follows a specific plan such as the construction of an edifice following one or more construction plans.
R10.2 <sub>akm</sub>	<i>followed general plan (was general plan of)</i>	E7 Activity	E55 Type	<i>Sub-property of P32</i> This property describes the relationship between an activity (E7) and a type (E55) of plan (C4 <sub>akm</sub> ) upon which said activity is based.

ID	Property Name	Domain	Range	Scope Note
R11 <sub>akm</sub>	<i>specifically planned (was specifically planned by)</i>	C6 <sub>akm</sub> Planning	E7 Activity	<p data-bbox="1180 296 1361 319"><i>Sub-property of P20</i></p> <p data-bbox="1180 328 1939 419">This property describes the relationship between a planning activity (C6<sub>akm</sub>) and a specific activity (E7) specifically planned, prepared or organized by said planning activity.</p>

## Native CRM Properties

The table lists all properties taken from the CRM (Crofts et al., 2006), whereby one property has been taken from FRBRoo (Bekiari et al., 2010), and reused in the AKM. Note that ranges and domains are given as designated in the CRM which means that classes may be listed which are not part of the AKM. In these cases one of the subclasses is part of the AKM. Please consult the documentation of the CRM for the complete class hierarchy.

Table 20 – Native CRM Properties

ID	Property Name	Domain	Range	Scope Note
P1	<i>is identified by (identifies)</i>	E1 CRM Entity	E41 Appellation	<p>This property describes the naming or identification of any real world item by a name or any other identifier.</p> <p>This property is intended for identifiers in general use, which form part of the world the model intends to describe, and not merely for internal database identifiers which are specific to a technical system, unless these latter also have a more general use outside the technical context. This property includes in particular identification by mathematical expressions such as coordinate systems used for the identification of instances of E53 Place. The property does not reveal anything about when, where and by whom this identifier was used. A more detailed representation can be made using the fully developed (i.e. indirect) path through E15 Identifier Assignment.</p>
P2	<i>has type (is type of)</i>	E1 CRM Entity	E55 Type	<p>This property allows sub typing of CRM entities – a form of specialisation - through the use of a terminological hierarchy, or thesaurus.</p> <p>The CRM is intended to focus on the high-level entities and relationships needed to describe data structures. Consequently, it does not specialise entities any further than is required for this immediate purpose. However, entities in the isA hierarchy of the CRM may be specialised into any number of sub entities, which can be defined in the E55 Type hierarchy. E51 Contact Point, for example, may be specialised into "e-mail address", "telephone number", "post office box", "URL" etc. none of which figures explicitly in the CRM hierarchy. Sub typing obviously requires consistency between the meaning of the terms assigned and the more general intent of the CRM entity in question.</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P4	<i>has time-span (is time-span of)</i>	E2 Temporal Entity	E52 Time-Span	<p>This property describes the temporal confinement of an instance of an E2 Temporal Entity.</p> <p>The related E52 Time-Span is understood as the real Time-Span during which the phenomena were active, which make up the temporal entity instance. It does not convey any other meaning than a positioning on the "time-line" of chronology. The Time-Span in turn is approximated by a set of dates (E61 Time Primitive). A temporal entity can have in reality only one Time-Span, but there may exist alternative opinions about it, which we would express by assigning multiple Time-Spans. Related temporal entities may share a Time-Span. Time-Spans may have completely unknown dates but other descriptions by which we can infer knowledge.</p>
P7	<i>took place at (witnessed)</i>	E4 Period	E53 Place	<p>This property describes the spatial location of an instance of E4 Period.</p> <p>The related E53 Place should be seen as an approximation of the geographical area within which the phenomena that characterise the period in question occurred. P7 took place at (witnessed) does not convey any meaning other than spatial positioning (generally on the surface of the earth). For example, the period "Révolution française" can be said to have taken place in "France", the "Victorian" period, may be said to have taken place in "Britain" and its colonies, as well as other parts of Europe and north America. A period can take place at multiple locations.</p>
P9	<i>consists of (forms part of)</i>	E4 Period	E4 Period	<p>This property describes the decomposition of an instance of E4 Period into discrete, subsidiary periods.</p> <p>The sub-periods into which the period is decomposed form a logical whole - although the entire picture may not be completely known - and the sub-periods are constitutive of the general period.</p>
P10	<i>falls within (contains)</i>	E4 Period	E4 Period	<p>This property describes an instance of E4 Period, which falls within the E53 Place and E52 Time-Span of another.</p> <p>The difference with P9 consists of (forms part of) is subtle. Unlike P9 consists of (forms part of), P10 falls within (contains) does not imply any logical connection between the two periods and it may refer to a period of a completely different type.</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P11	<i>had participant (participated in)</i>	E5 Event	E39 Actor	<p>This property describes the active or passive participation of instances of E39 Actors in an E5 Event.</p> <p>It connects the life-line of the related E39 Actor with the E53 Place and E50 Date of the event. The property implies that the Actor was involved in the event but does not imply any causal relationship. The subject of a portrait can be said to have participated in the creation of the portrait.</p>
P14	<i>carried out by (performed)</i>	E7 Activity	E39 Actor	<p>This property describes the active participation of an E39 Actor in an E7 Activity. It implies causal or legal responsibility. The P14.1 in the role of property of the property allows the nature of an Actor's participation to be specified.</p>
P15	<i>was influenced by (influenced)</i>	E7 Activity	E1 CRM Entity	<p>This is a high level property, which captures the relationship between an E7 Activity and anything that may have had some bearing upon it.</p> <p>The property has more specific sub properties.</p>
P17	<i>was motivated by (motivated)</i>	E7 Activity	E1 CRM Entity	<p>This property describes an item or items that are regarded as a reason for carrying out the E7 Activity.</p> <p>For example, the discovery of a large hoard of treasure may call for a celebration, an order from head quarters can start a military manoeuvre.</p>
P21	<i>had general purpose (was purpose of)</i>	E7 Activity	E55 Type	<p>This property describes an intentional relationship between an E7 Activity and some general goal or purpose.</p> <p>This may involve activities intended as preparation for some type of activity or event. P21had general purpose (was purpose of) differs from P20 had specific purpose (was purpose of) in that no occurrence of an event is implied as the purpose.</p>
P26	<i>moved to (was destination of)</i>	E9 Move	E53 Place	<p>This property identifies the destination of a E9 Move.</p> <p>A move will be linked to a destination, such as the move of an artefact from storage to display. A move may be linked to many terminal instances of E53 Places. In this case the move describes a distribution of a set of objects. The area of the move includes the origin, route and destination.</p>
P26	<i>moved from (was origin of)</i>	E9 Move	E53 Place	<p>This property identifies the starting E53 Place of an E9 Move.</p> <p>A move will be linked to an origin, such as the move of an artefact from storage to display. A move may be linked to many origins. In this case the move describes the picking up of a set of objects. The area of the move includes the origin, route and destination.</p>



<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P28	<i>custody surrendered by (surrendered custody through)</i>	E10 Transfer of Custody	E39 Actor	<p>This property identifies the E39 Actor or Actors who surrender custody of an instance of E18 Physical Thing in an E10 Transfer of Custody activity.</p> <p>The property will typically describe an Actor surrendering custody of an object when it is handed over to someone else's care. On occasion, physical custody may be surrendered involuntarily - through accident, loss or theft. In reality, custody is either transferred to someone or from someone, or both.</p>
P29	<i>custody received by (received custody through)</i>	E10 Transfer of Custody	E39 Actor	<p>This property identifies the E39 Actor or Actors who receive custody of an instance of E18 Physical Thing in an E10 Transfer of Custody activity.</p> <p>The property will typically describe Actors receiving custody of an object when it is handed over from another Actor's care. On occasion, physical custody may be received involuntarily or illegally - through accident, unsolicited donation, or theft. In reality, custody is either transferred to someone or from someone, or both.</p>
P30	<i>transferred custody of (custody transferred through)</i>	E10 Transfer of Custody	E18 Physical Thing	<p>This property identifies an item or items of E18 Physical Thing concerned in an E10 Transfer of Custody activity.</p> <p>The property will typically describe the object that is handed over by an E39 Actor to another Actor's custody. On occasion, physical custody may be transferred involuntarily or illegally - through accident, unsolicited donation, or theft.</p>
P43	<i>has dimension (is dimension of)</i>	E70 Thing	E54 Dimension	<p>This property records a E54 Dimension of some E70 Thing. It is a shortcut of the more fully developed path from E70 Thing through P39 measured (was measured by), E16 Measurement P40 observed dimension (was observed in) to E54 Dimension. It offers no information about how and when an E54 Dimension was established, nor by whom.</p> <p>An instance of E54 Dimension is specific to an instance of E70 Thing.</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P46	<i>is composed of (forms part of)</i>	E18 Physical Thing	E18 Physical Thing	<p>This property allows instances of E18 Physical Thing to be analysed into component elements.</p> <p>Component elements, since they are themselves instances of E18 Physical Thing, may be further analysed into sub-components, thereby creating a hierarchy of part decomposition. An instance of E18 Physical Thing may be shared between multiple wholes, for example two buildings may share a common wall.</p> <p>This property is intended to describe specific components that are individually documented, rather than general aspects. Overall descriptions of the structure of an instance of E18 Physical Thing are captured by the P3 has note property.</p> <p>The instances of E57 Materials of which an item of E18 Physical Thing is composed should be documented using P45 consists of (is incorporated in).</p>
P48	<i>has preferred identifier (is preferred identifier of)</i>	E1 CRM Entity	E42 Identifier	<p>This property records the preferred E42 Identifier that was used to identify an instance of E1 CRM Entity at the time this property was recorded.</p> <p>More than one preferred identifier may have been assigned to an item over time. Use of this property requires an external mechanism for assigning temporal validity to the respective CRM instance.</p> <p>P48 has preferred identifier (is preferred identifier of), is a shortcut for the path from E1 CRM Entity through P140 assigned attribute to (was attributed by), E15 Identifier Assignment, P37 assigned (was assigned by) to E42 Identifier. The fact that an identifier is a preferred one for an organisation can be better expressed in a context independent form by assigning a suitable E55 Type to the respective instance of E15 Identifier Assignment using the P2 has type property.</p>
P49	<i>has former or current keeper (is former or current keeper of)</i>	E18 Physical Thing	E39 Actor	<p>This property identifies the E39 Actor or Actors who have or have had custody of an instance of E18 Physical Thing at some time.</p> <p>The distinction with P50 has current keeper (is current keeper of) is that P49 has former or current keeper (is former or current keeper of) leaves open the question as to whether the specified keepers are current.</p> <p>P49 has former or current keeper (is former or current keeper of) is a shortcut for the more detailed path from E18 Physical Thing through P30 transferred custody of (custody transferred through), E10 Transfer of Custody, P28 custody surrendered by (surrendered custody through) or P29 custody received by (received custody through) to E39 Actor.</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P50	<i>has current keeper (is current keeper of)</i>	E18 Physical Thing	E39 Actor	<p>This property identifies the E39 Actor or Actors who had custody of an instance of E18 Physical Thing at the time this property was recorded.</p> <p>P50 has current keeper (is current keeper of) is a shortcut for the more detailed path from E18 Physical Thing through P30 transferred custody of (custody transferred through), E10 Transfer of Custody, P29 custody received by (received custody through) to E39 Actor.</p>
P51	<i>has former or current owner (is former or current owner of)</i>	E18 Physical Thing	E39 Actor	<p>This property identifies the E39 Actor that is or has been the legal owner (i.e. title holder) of an instance of E18 Physical Thing at some time.</p> <p>The distinction with P52 has current owner (is current owner of) is that P51 has former or current owner (is former or current owner of) does not indicate whether the specified owners are current. P51 has former or current owner (is former or current owner of) is a shortcut for the more detailed path from E18 Physical Thing through P24 transferred title of (changed ownership through), E8 Acquisition, P23 transferred title from (surrendered title through), or P22 transferred title to (acquired title through) to E39 Actor.</p>
P54	<i>has current permanent location (is current permanent location of)</i>	E19 Physical Object	E53 Place	<p>This property records the foreseen permanent location of an instance of E19 Physical Object at the time this property was recorded.</p> <p>P54 has current permanent location (is current permanent location of) is similar to P55 has current location (currently holds). However, it indicates the E53 Place currently reserved for an object, such as the permanent storage location or a permanent exhibit location. The object may be temporarily removed from the permanent location, for example when used in temporary exhibitions or loaned to another institution. The object may never actually be located at its permanent location.</p>
P55	<i>has current location (currently holds)</i>	E19 Physical Object	E53 Place	<p>This property records the location of an E19 Physical Object at the time the property was recorded.</p> <p>This property is a specialisation of P53 has former or current location (is former or current location of). It indicates that the E53 Place associated with the E19 Physical Object is the current location of the object. The property does not allow any indication of how long the Object has been at the current location. P55 has current location (currently holds) is a shortcut. A more detailed representation can make use of the fully developed (i.e. indirect) path from E19 Physical Object through P25 moved (moved by), E9 Move P26 moved to (was destination of) to E53 Place if and only if this Move is the most recent.</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P67	<i>refers to (is referred to by)</i>	E89 Propositional Object	E1 CRM Entity	This property documents that an E89 Propositional Object makes a statement about an instance of E1 CRM Entity. P67 refers to (is referred to by) has the P67.1 has type link to an instance of E55 Type. This is intended to allow a more detailed description of the type of reference. This differs from P129 is about (is subject of), which describes the primary subject or subjects of the E89 Propositional Object.
P69	<i>is associated with</i>	E29 Design or Procedure	E29 Design or Procedure	This symmetric property describes the association of an E29 Design or Procedure with other Designs or Procedures. Any instance of E29 Design or Procedure may be associated with other designs or procedures. The nature of the association may be whole-part, sequence, prerequisite etc. The property is assumed to be entirely reciprocal.
P70	<i>documents (is documented in)</i>	E31 Document	E1 CRM Entity	This property describes the CRM Entities documented by instances of E31 Document. Documents may describe any conceivable entity, hence the link to the highest-level entity in the CRM hierarchy. This property is intended for cases where a reference is regarded as being of a documentary character, in the scholarly or scientific sense.
P74	<i>has current or former residence (is current or former resident of)</i>	E39 Actor	E53 Place	This property describes the current or former E53 Place of residence of an E39 Actor. The residence may be either the Place where the Actor resides, or a legally registered address of any kind.
P94	<i>has created (was created by)</i>	E65 Creation	E28 Conceptual Object	This property allows a conceptual E65 Creation to be linked to the E28 Conceptual Object created by it. It represents the act of conceiving the intellectual content of the E28 Conceptual Object. It does not represent the act of creating the first physical carrier of the E28 Conceptual Object. As an example, this is the composition of a poem, not its commitment to paper.
P95	<i>has formed (was formed by)</i>	E66 Formation	E74 Group	This property links the founding or E66 Formation for an E74 Group with the Group itself.
P98	<i>brought into life (was born)</i>	E67 Birth	E21 Person	This property links an E67Birth event to an E21 Person in the role of offspring. Twins, triplets etc. are brought into life by the same Birth event. This is not intended for use with general Natural History material, only people. There is no explicit method for modelling conception and gestation except by using extensions.

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P99	<i>dissolved (was dissolved by)</i>	E68 Dissolution	E74 Group	This property links the disbanding or E68 Dissolution of an E74 Group to the Group itself.
P100	<i>was death of (died in)</i>	E69 Death	E21 Person	This property links an E69 Death event to the E21 Person that died. A Death event may involve multiple people, for example in the case of a battle or disaster. This is not intended for use with general Natural History material, only people.
P101	<i>had as general use (was use of)</i>	E70 Thing	E55 Type	This property links an instance of E70 Thing to an E55 Type of usage. It allows the generic link between things, both physical and immaterial, to methods and techniques of use. Thus it can be asserted that a baseball bat had a general use for sport and a specific use for threatening people during the Great Train Robbery.
P102	<i>has title (is title of)</i>	E71 Man-Made Thing	E35 Title	This property describes the E35 Title applied to an instance of E71 Man-Made Thing. The E55 Type of Title is assigned in a sub property. The P102.1 has type property of the P102 has title (is title of) property enables the relationship between the Title and the thing to be further clarified, for example, if the Title was a given Title, a supplied Title etc. It allows any man-made material or immaterial thing to be given a Title. It is possible to imagine a Title being created without a specific object in mind.
P107	<i>has current or former member (is current or former member of)</i>	E74 Group	E39 Actor	This property relates an E39 Actor to the E74 Group of which he or she is a member. Groups, Legal Bodies and Persons, may all be members of Groups. A Group necessarily consists of more than one member. This property is a shortcut of the more fully developed path from E74 Group through P144 joined with (gained member by), E85 Joining, P143 joined (was joined by) to E39 Actor The property P107.1 kind of member can be used to specify the type of membership or the role the member has in the group.
P108	<i>has produced (was produced by)</i>	E12 Production	E24 Physical Man-Made Thing	This property identifies the E24 Physical Man-Made Thing that came into existence as a result of an E12 Production. The identity of an instance of E24 Physical Man-Made Thing is not defined by its matter, but by its existence as a subject of documentation. An E12 Production can result in the creation of multiple instances of E24 Physical Man-Made Thing.

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P109	<i>has current or former curator (is current or former curator of)</i>	E78 Collection	E39 Actor	<p>This property identifies the E39 Actor or Actors who assume or have assumed overall curatorial responsibility for an E78 Collection.</p> <p>This property is effectively a short-cut. It does not allow a history of curation to be recorded. This would require use of an Event assigning responsibility for a Collection to a curator.</p>
P128	<i>carries (is carried by)</i>	E24 Physical Man-Made Thing	E73 Information Object	<p>This property identifies an E73 Information Object carried by an instance of E24 Physical Man-Made Thing.</p> <p>In general this would be an E84 Information Carrier P65 shows visual item (is shown by) is a specialisation of P128 carries (is carried by) which should be used for carrying visual items.</p>
P129	<i>is about (is subject of)</i>	E89 Propositional Object	E1 CRM Entity	<p>This property documents that an E89 Propositional Object has as subject an instance of E1 CRM Entity.</p> <p>This differs from P67 refers to (is referred to by), which refers to an E1 CRM Entity, in that it describes the primary subject or subjects of an E89 Propositional Object.</p>
P134	<i>continued (was continued by)</i>	E7 Activity	E7 Activity	<p>This property allows two activities to be related where the domain is considered as an intentional continuation of the range.</p> <p>Used multiple times, this allows a chain of related activities to be created which follow each other in sequence.</p>
P143	<i>joined (was joined by)</i>	E85 Joining	E39 Actor	<p>This property identifies the instance of E39 Actor that becomes member of a E74 Group in an E85 Joining.</p> <p>Joining events allow for describing people becoming members of a group with a more detailed path from E74 Group through P144 joined with (gained member by), E85 Joining, P143 joined (was joined by) to E39 Actor, compared to the shortcut offered by P107 has current or former member (is current or former member of).</p>

<b>ID</b>	<b>Property Name</b>	<b>Domain</b>	<b>Range</b>	<b>Scope Note</b>
P144	<i>joined with (gained member by)</i>	E85 Joining	E74 Group	<p>This property identifies the instance of E74 Group of which an instance of E39 Actor becomes a member through an instance of E85 Joining.</p> <p>Although a Joining activity normally concerns only one instance of E74 Group, it is possible to imagine circumstances under which becoming member of one Group implies becoming member of another Group as well.</p> <p>Joining events allow for describing people becoming members of a group with a more detailed path from E74 Group through P144 joined with (gained member by), E85 Joining, P143 joined (was joined by) to E39 Actor, compared to the shortcut offered by P107 has current or former member (is current or former member of). The property P144.1 kind of member can be used to specify the type of membership or the role the member has in the group.</p>
P145	<i>separated (left by)</i>	E86 Leaving	E39 Actor	<p>This property identifies the instance of E39 Actor that leaves an instance of E74 Group through an instance of E86 Leaving.</p>
P146	<i>separated from (lost member by)</i>	E86 Leaving	E74 Group	<p>This property identifies the instance of E74 Group an instance of E39 Actor leaves through an instance of E86 Leaving.</p> <p>Although a Leaving activity normally concerns only one instance of E74 Group, it is possible to imagine circumstances under which leaving one E74 Group implies leaving another E74 Group as well.</p>
P148	<i>has component (is component of)</i>	E89 Propositional Object	E89 Propositional Object	<p>This property associates an instance of E89 Propositional Object with a structural part of it that is by itself an instance of E89 Propositional Object.</p>
R29	<i>reproduced (was reproduced by)</i>	F32 Carrier Production Event	F5 Item	<p>Sub-Property of P108 has produced (was produced by). This property associates an instance of F32 Carrier Production Event with any one of the produced items (i.e., the instances of F5 Item).</p>

## References

- Margaret O'Neill Adams. Analyzing Archives and Finding Facts: Use and Users of Digital Data Records. *Archival Science*, 7(1):21–36, 2007.
- Vladimir Alexiev. Implementing CIDOC CRM Search Based on Fundamental Relations and OWLIM Rules, 2011.
- Burt Altman and John R. Nemmers. The Usability of On-line Archival Resources: The POLARIS Project Finding Aid. *The American Archivist*, 64(1):121–131, 2001.
- Ian G. Anderson. Are You Being Served? Historians and the Search for Primary Sources. *Archivaria*, 58(Fall):81–129, 2004.
- Ian G. Anderson. Necessary but Not Sufficient: Modelling Online Archive Development in the UK. *D-Lib Magazine*, 14(1/2), 2008.
- Kimberly Anderson. The Footprint and the Stepping Foot: Archival Records, Evidence, and Time. *Archival Science*, 13(4):349–371, 2013.
- Grigoris Antoniou, Paul Groth, Frank van Harmelen, and Rinke Hoekstra. *A Semantic Web Primer*. Cooperative information systems. The MIT Press, Cambridge, MA, 2012.
- Steve Bailey. Taking the Road Less Travelled By: The Future of the Archive and Records Management Profession in the Digital Age. *Journal of the Society of Archivists*, 28(2):117–124, 2007.
- Kyle Banerjee and Max Johnson. Improving Access to Archival Collections with Automated Entity Extraction. *code4lib*, (29), 2015.
- Marcia J. Bates. Berrypicking. In Karen E. Fisher, Sanda Erdelez, and Lynne McKechnie, editors, *Theories of Information Behavior*, ASIST monograph series, pages 58–62. Information Today, Medford, 2009.
- Marcia J. Bates. Information Behavior. In Marcia J. Bates and Mary Niles Maack, editors, *Encyclopedia of Library and Information Sciences*, pages 2381–2391. CRC Press, New York, 2010.
- David Bearman. User Presentation Language in Archives. *Archives & Museum Informatics*, 3(4): 3–7, 1989/90.
- Diane L. Beattie. An Archival User Study: Researchers in the Field of Women's History. *Archivaria*, 29(Winter):33–50, 1989/90.
- Carl L. Becker. What are Historical Facts? *The Western Political Quarterly*, 8(3):327–340, 1955.
- Chryssoula Bekiari, Martin Doerr, and Patrick Le Boeuf. FRBR Object-Oriented Definition and Mapping to FRBRER: Version 1.0.1, 2010.



- Nicholas J. Belkin. Anomalous States of Knowledge as a Basis for Information Retrieval. *Canadian Journal of Information and Library Science*, 5:133–143, 1980.
- Ernst Bernheim. *Lehrbuch der historischen Methode und der Geschichtsphilosophie: Mit Nachweis der wichtigsten Quellen und Hilfsmittel zum Studium der Geschichte*. Duncker & Humblot, Leipzig, 3. und 4. völlig neu bearb. und verm. aufl. edition, 1903.
- Marc Bloch and Jacques Le Goff. *Apologie der Geschichtswissenschaft oder der Beruf des Historikers*. Klett-Cotta, Stuttgart, 2 edition, 2008.
- Onno Boonstra, Leen Breure, and Peter Doorn. *Past, Present and Future of Historical Information Science*. Niwi-Knaw, Amsterdam, 2004.
- Egon Boshof, Kurt Düwell, and Hans Kloft. *Grundlagen des Studiums der Geschichte: Eine Einführung*. Böhlau-Studien-Bücher. Böhlau, Köln and Weimar and Wien, 5., durchges. aufl. edition, 1997.
- Lina Bountouri and Manolis Gergatsoulis. Mapping Encoded Archival Description to CIDOC CRM. In Christos Papatheodorou and Manolis Gergatsoulis, editors, *Proceedings of the First Workshop on Digital Information Management*, pages 8–25, Corfu, 2011.
- Françoise Bourdon, Christina Hengel-Dittrich, Olga Lavrenova, Andrew MacEwan, Eeva Murto-maa, Glenn E. Patton, Henry Snyder, Barbara Tillett, Hartmut Walravens, and Mirna Willer. Functional Requirements for Authority Data (FRAD): A Conceptual Model: Final Report, December 2008, 2013.
- Ronald J. Brachman. What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *Computer*, 16(10):30–36, 1983.
- Sabine Brenner-Wilczek, Gertrude Cepl-Kaufmann, and Max Plassmann. *Einführung in die moderne Archivarbeit*. WBG (Wissenschaftliche Buchgesellschaft), Darmstadt, 2006.
- Kate Byrne. Putting Hybrid Cultural Data on the Semantic Web. *Journal of Digital Information*, 10(6), 2009.
- Kate Byrne and Ewan Klein. Automatic Extraction of Archaeological Events from Text. In Bernard Fischer, J. Webb Crawford, and David Koller, editors, *Making History Interactive: CAA 2009*, BAR International Series, pages 48–56, Oxford, 2009. Archaeopress.
- Diego Calvanese, Giuseppe de Giacomo, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati. Description Logic Framework for Information Integration. In Anthony G. Cohn, Lenhart K. Schubert, and Stuart C. Shapiro, editors, *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning*, pages 2–13. Morgan Kaufmann, 1998.
- Ruth Carter and Thomas Frusciano. Online Finding Aids and Users of Archives: Continuing the Dialog. *Journal of Archival Organization*, 2(3):1–5, 2004.

- Donald Owen Case. The Collection and Use of Information by Some American Historians: A Study of Motives and Methods. *Library Quarterly*, 61(1):61–82, 1991.
- Donald Owen Case. *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Library and Information Science. Elsevier, Amsterdam, 2. ed. edition, 2008.
- Joyce Celeste Chapman. Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids. 8(1):4–30, 2010.
- Hsin-liang Chen. An Analysis of Image Queries in the Field of Art History. *Journal of the American Society for Information Science and Technology*, 52(3):260–273, 2001.
- Youngok Choi and Edie M. Rasmussen. Searching for Images: The Analysis of Users' Queries for Image Retrieval in American History. *Journal of the American Society for Information Science and Technology*, 54(6):498–511, 2003.
- Paul Christianson. Patterns of Historical Interpretation. In Dussen, W. Jan van der and Lionel Rubinoff, editors, *Objectivity, Method, and Point of View*, Philosophy of History and Culture, pages 47–71. Brill, Leiden, 1991.
- Malcolm Clark, Yunhyong Kim, Udo Kruschwitz, Dawei Song, Dyaa Albakour, Stephen Dignum, Ulises Cerviño Beresi, Maria Fasli, and Anne de Roeck. Automatically Structuring Domain Knowledge from Text: An Overview of Current Research. *Information Processing & Management*, 48(3):552–568, 2012.
- Lisa R. Coats. Users of EAD Finding Aids: Who Are They and Are They Satisfied? *Journal of Archival Organization*, 2(3):25–39, 2004.
- Robin G. Collingwood. The Idea of History. In Leonard M. Marsak, editor, *The Nature of Historical Inquiry*, pages 11–20. Holt Rinehart and Winston, New York, 1970.
- Karen Collins. Providing Subject Access to Images: A Study of User Queries. *The American Archivist*, 61(1):36–55, 1998.
- Paul Conway. Facts and Frameworks: An Approach to Studying the Users of Archives. *The American Archivist*, 49(4):393–407, 1986.
- Paul Conway. *Partners in Research: Improving Access to the Nation's Archive*, 1994.
- Terry Cook. From Information to Knowledge: An Intellectual Paradigm for Archives. *Archivaria*, 19(Winter):28–49, 1984.
- Terry Cook. The Archive(s) Is a Foreign Country: Historians, Archivists, and the Changing Archival Landscape. *The American Archivist*, 74(Fall/Winter):600–632, 2011.
- Richard J. Cox. Researching Archival Reference as an Information Function: Observations on Needs and Opportunities. *Reference Quarterly*, 31(3):387–397, 1992.

- Richard J. Cox. Revisiting the Archival Finding Aid. *Journal of Archival Organization*, 5(4):5–32, 2008.
- Barbara Craig. Old Myths in New Clothes: Expectations of Archives Users. (45):118–126, 1998.
- Barbara Craig. Perimeters with Fences? Or Thresholds with Doors? Two Views of a Border. *The American Archivist*, 66(1):96–101, 2003.
- Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. Definition of the CIDOC Conceptual Reference Model: Produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group: Version 4.2.1, October 2006, 2006.
- Kate Cruikshank, Caroline Daniels, Dennis Meissner, Naomi L. Nelson, and Mark Shelstad. How Do We Show You What We've Got? Access to Archival Collections in the Digital Age. *Journal of the Association for History and Computing*, 8(2), 2005.
- Wolfgang Degen, Barbara Heller, Heinrich Herre, and Barry Smith. GOL: Toward an Axiomatized Upper-level Ontology. In Nicola Guarino, Barry Smith, and Christopher Welty, editors, *Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001*, pages 34–46, New York, 2001. ACM.
- Martin Doerr. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, 2003.
- Martin Doerr and Gerald Hiebel. CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement: Technical Report: ICS-FORTH/TR-435, April 2013, 2013.
- Martin Doerr and Dolores Iorizzo. The Dream of a Global Knowledge Network: A New Approach. *Journal on Computing and Cultural Heritage*, 1(1):5:1–5:23, 2008.
- Martin Doerr, Nick Crofts, and Tony Gill. The CIDOC Conceptual Reference Model: A Standard for Communicating Cultural Contents. *Cultivate Interactive*, (9), 2003.
- Martin Doerr, Christian-Emil Ore, and Stephen Stead. The CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing: ER 2007 Tutorial. In John C. Grundy, Sven Hartmann, Alberto H. F. Laender, Leszek A. Maciaszek, and John F. Roddick, editors, *Challenges in Conceptual Modelling: ER 2007*, Conferences in Research and Practice in Information Technology, pages 51–56, Darlinghurst, Australia, 2007. Australian Computer Society.
- Martin Doerr, Athina Kritsotaki, and Stephen Stead. Which Period is it? A Methodology to Create Thesauri of Historical Periods. In Franco Niccolucci and Sorin Hermon, editors, *Beyond the Artifact: Proceedings of CAA2004*, pages 70–75, Budapest, 2010a. Archeolingua.

- Martin Doerr, Dimitris Plexousakis, Katerina Kopaka, and Chryssoula Bekiari. Supporting Chronological Reasoning in Archaeology. In Franco Niccolucci and Sorin Hermon, editors, *Beyond the Artifact: Proceedings of CAA2004*, pages 13–17, Budapest, 2010b. Archeolingua.
- Martin Doerr, Athina Kritsotaki, and Katerina Boutsika. Factual Argumentation: A Core Model for Assertions Making. *Journal on Computing and Cultural Heritage*, 3(3):8:1–8:34, 2011.
- Martin Doerr, Athina Kritsotaki, Yannis Rousakis, Gerald Hiebel, and Maria Theodoridou. CRMsci: The Scientific Observation Model: An Extension of CIDOC-CRM to Support Scientific Observation: Version 1.2 (draft), 2014.
- Jackie M. Dooley. Subject Indexing in Context. *The American Archivist*, 55(1):344–354, 1992.
- Elizabeth H. Dow. Encoded Archival Description as a Halfway Technology. *Journal of Archival Organization*, 7(3):108–115, 2009.
- Lawrence Dowler. The Role of Use in Defining Archival Practice and Principles: A Research Agenda for the Availability and Use of Records. 51(1):74–86, 1988.
- Johann Gustav Droysen. *Grundriss der Historik*. Veit, Leipzig, 1868.
- Jean Dryden. Do We Care What Users Want? Evaluating User Satisfaction and the LibQUAL+™ Experience. *Journal of Archival Organization*, 2(4):83–88, 2005.
- Wendy Duff. Understanding the Information-seeking Behaviour of Archival Researchers in the Digital Age: Path, Processes and Preferences. In Peter Berninger, Frank Brady, Hans Hofmann, and Jef Schram, editors, *Proceedings of the DLM-Forum 2002: Best Practices and Solutions*, Insar, pages 331–339, Brüssel, 2002.
- Wendy Duff and Allyson Fox. ‘You’re a Guide Rather Than an Expert’: Archival Reference from an Archivist’s Point of View. *Journal of the Society of Archivists*, 27(2):129–153, 2006.
- Wendy M. Duff and Catherine A. Johnson. A Virtual Expression of Need: An Analysis of E-mail Reference Questions. *The American Archivist*, 64(1):43–60, 2001.
- Wendy M. Duff and Catherine A. Johnson. Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. *Library Quarterly*, 72(4):472–496, 2002.
- Wendy M. Duff and Catherine A. Johnson. Where Is the List with All the Names? Information-Seeking Behavior of Genealogists. *The American Archivist*, 66(1):79–95, 2003.
- Wendy M. Duff and Penka Stoyanova. Transforming the Crazy Quilt: Archival Displays from a Users’ Point of View. *Archivaria*, 45:44–79, 1998.
- Wendy M. Duff, Barbara Craig, and Joan Cherry. Finding and Using Archival Resources: A Cross-Canada Survey of Historians Studying Canadian History. *Archivaria*, 58:51–80, 2004a.

- Wendy M. Duff, Barbara Craig, and Joan Cherry. Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age. *The Public Historian*, 26(2):7–22, 2004b.
- Wendy M. Duff, Jean Dryden, Carry Limkilde, Joan Cherry, and Ellie Bogomazova. Archivists' Views on User-based Evaluation: Benefits, Barriers, and Requirements. *The American Archivist*, 71(Spring/Summer):144–166, 2008.
- Garett O. Dworman, Steven O. Kimbrough, and Chuck Patch. On Pattern-Directed Search of Archives and Collections. *Journal of the American Society for Information Science and Technology*, 51(1):14–23, 2000.
- Matthew Young Eidson. Describing Anything That Walks: The Problem Behind the Problem of EAD. *Journal of Archival Organization*, 1(4):5–28, 2002.
- David Ellis. Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach. *Library Quarterly*, 63(4):469–488, 1993.
- David Ellis. Ellis's Model of Information-Seeking Behavior. In Karen E. Fisher, Sanda Erdelez, and Lynne McKechnie, editors, *Theories of Information Behavior*, ASIST monograph series, pages 138–142. Information Today, Medford, 2009.
- Max J. Evans. Archives of the People, by the People, for the People. *The American Archivist*, 70 (Fall/Winter):387–400, 2007.
- Karl-Georg Faber. *Theorie der Geschichtswissenschaft*, volume 78 of *Beck'sche Schwarze Reihe*. Beck, München, 5., erw. Aufl. edition, 1982.
- Gilles Fauconnier and Mark Turner. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York, 2003.
- Kathleen Fear. User Understanding of Metadata in Digital Image Collections: Or, What Exactly Do You Mean by 'Coverage'? *The American Archivist*, 73(1):26–60, 2009.
- Edward A. Feigenbaum and Pamela McCorduck. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. Addison-Wesley, Reading, 2 edition, 1983.
- Candida Fenton. Use of Controlled Vocabulary and Thesauri in UK Online Finding Aids. *Journal of the Society of Archivists*, 31(2):187–205, 2010.
- Elsie Freeman Finch, editor. *Advocating Archives: An Introduction to Public Relations for Archivists*. Society of American Archivists and Scarecrow Press, Metuchen, N.J, 1994.
- Michael J. Fox. Stargazing: Locating EAD in the Descriptive Firmament. In Daniel V. Pitti and Wendy M. Duff, editors, *Encoded Archival Description on the Internet*, pages 61–74. Haworth Information Press, Binghamton NY, 2001.

- Michael J. Fox, Peter L. Wilkerson, and Susanne R. Warren. *Introduction to Archival Organization and Description*. Getty Information Institute, Los Angeles, 1998.
- Eckhart G. Franz. *Einführung in die Archivkunde*. Die Geschichtswissenschaft. WBG (Wissenschaftliche Buchgesellschaft), Darmstadt, 6 edition, 2004.
- Elsie T. Freeman. In the Eye of the Beholder: Archives Administration from the User's Point of View. *The American Archivist*, 47(2):111–123, 1984.
- Thomas J. Frusciano. "Ten Years After": The Next Wave of EAD Implementation. *Journal of Archival Organization*, 5(3):1–8, 2008a.
- Thomas J. Frusciano. The Role of Technology in Users' Expectations. *Journal of Archival Organization*, 6(4):213–215, 2008b.
- Louise Gagnon-Arguin. Les Questions de Recherche Comme Matériau d'Études des Usagers en Vue du Traitement des Archives. 46(Fall):86–102, 1998.
- Jean-Claude Gardin. Archaeological Discourse, Conceptual Modelling and Digitalisation: An Interim Report of the Logician Program. In Martin Doerr and Apostolos Sarris, editors, *The Digital Heritage of Archaeology: CAA 2002*, pages 5–11, Athen, 2003. Archive of Monuments and Publications, Hellenic Ministry of Culture.
- Gilbert J. Garraghan. *A Guide to Historical Method*. Fordham University Press, New York, 4 edition, 1957.
- Dedre Gentner and Albert L. Stevens, editors. *Mental Models*. Cognitive science. Erlbaum and Taylor and Francis, Hillsdale, N.J, 1983.
- Anne J. Gilliland-Swetland. Popularizing the Finding Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups. In Daniel V. Pitti and Wendy M. Duff, editors, *Encoded Archival Description on the Internet*, pages 199–227. Haworth Information Press, Binghamton NY, 2001.
- Victor Gray. 'Who's that Knocking on Our Door? Archives, Outreach and Community. *Journal of the Society of Archivists*, 29(1):1–8, 2008.
- Jacques Grimard. Managing the Long-term Preservation of Electronic Archives or Preserving the Medium and the Message. *Archivaria*, (59):153–167, 2005.
- Denis J. Grogan. *Practical Reference Work*. Library Association Publishing, London, 2 edition, 1992.
- Thomas R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.

- Tom Gruber. Ontology, 2007. URL <http://tomgruber.org/writing/ontology-definition-2007.htm>.
- Nicola Guarino. Formal Ontology and Information Systems. In Nicola Guarino, editor, *Formal Ontology in Information Systems: FOIS '98*, Frontiers in artificial intelligence and applications, pages 3–15, Amsterdam, 1998. IOS-Press.
- Nicola Guarino and Giarretta Pierdaniele. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. J. I. Mars, editor, *Towards Very Large Knowledge Bases*, pages 25–32, Amsterdam, 1995. IOS Press.
- Susan Hamburger. How Researchers Search for Manuscript and Archival Collections. *Journal of Archival Organization*, 2(1-2):79–102, 2004.
- Carolyn Harris. Archives Users in the Digital Era: A Review of Current Research Trends. *Dalhousie Journal of Interdisciplinary Management*, 1(Spring), 2005.
- Kent M. Haworth. Archival Description: Content and Context in Search of Structure. In Daniel V. Pitti and Wendy M. Duff, editors, *Encoded Archival Description on the Internet*, pages 7–26. Haworth Information Press, Binghamton NY, 2001.
- Steffen Henniecke. Leveraging EAD in a Semantic Web Environment to Enhance the Discovery Experience for the User in Digital Archives. In Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt, editors, *Research and Advanced Technology for Digital Libraries: TPDFL 2011, Proceedings*, Lecture Notes in Computer Science, pages 511–514, Berlin, 2011a. Springer.
- Steffen Henniecke. Leveraging EAD in a Semantic Web Environment to Enhance the Discovery Experience for the User in Digital Archives. In Luanne Freund, Edie Rasmussen, and Kazunari Sugiyama, editors, *Proceedings of the Doctoral Consortium*, pages 68–73, 2011b.
- Steffen Henniecke. Representation of Archival User Needs using CIDOC CRM. In Vladimir Alexiev, Vladimir Ivanov, and Maurice Grinberg, editors, *Practical Experiences with CIDOC CRM and its Extensions: CRMEX 2013*, CEUR Workshop Proceedings, pages 48–60, 2013.
- Steffen Henniecke. Modeling Patterns in Written Natural Language Questions to Archives. In Udo Kruschwitz, Frank Hopfgartner, and Cathal Gurrin, editors, *MindTheGap 2014*, CEUR Workshop Proceedings, pages 19–24, 2014.
- Thomas H. Hickerson. Ten Challenges for the Archival Profession. *The American Archivist*, 64 (Spring/Summer):6–16, 2001.
- Amanda Hill. Serving the Invisible Researcher: Meeting the Needs of Online Users. *Journal of the Society of Archivists*, 25(2):139–148, 2004.

- Christina J. Hostetter. Online Finding Aids: Are They Practical? *Journal of Archival Organization*, 2(1-2):117–145, 2004.
- Rachael Hu. Methods to Tame the Madness: A Practitioner’s Guide to User Assessment Techniques for Online Finding Aid and Website Design. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage*, 13(2):175–190, 2012.
- Noah G. Huffman. *Search Features and Other Characteristics of XML Retrieval Systems for EAD Finding Aids: A Content Analysis*. Master Thesis, University of North Carolina, Chapel Hill, North Carolina, 2008.
- Tim Hutchinson. Strategies for Searching Online Finding Aids: A Retrieval Experiment. *Archivaria*, 44:72–101, 1997.
- Nancy Ide and David Woolner. Exploiting Semantic Web Technologies for Intelligent Access to Historical Documents. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, and Raquel Silva, editors, *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, pages 2177–2180, Paris, 2004. European Language Resources Association (ELRA).
- Peter Ingwersen and Kalervo Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*, volume 18 of *Kluwer international series on information retrieval*. Springer, Dordrecht, 2010.
- International Council of Archives. *ISAD (G): General International Standard Archival Description: Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999*. ICA Standards. International Council of Archives, Ottawa, 2 edition, 2000.
- International Council of Archives. *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families: Adopted Canberra, Australia, 27 - 30. October 2003*, volume 2004 of *ICA Standards*. International Council of Archives, Paris, 2 edition, 2004.
- International Council of Archives. *ISDF: International Standard for Describing Functions: Developed by the Committee on Best Practices and Standards Dresden, Germany, 2-4 May 2007*. ICA Standards. International Council of Archives, Paris, 2007.
- International Council of Archives. *ISDIAH: International Standard for Describing Institutions with Archival Holdings: Developed by the Committee on Best Practices and Standards London, United Kingdom, 10-11 March 2008*. ICA Standards. International Council of Archives, Paris, 2008.
- John P. A. Ioannidis, Marcus R. Munafò, Paolo Fusar-Poli, Brian A. Nosek, and Sean P. David. Publication and Other Reporting Biases in Cognitive Sciences: Detection, Prevalence, and Prevention. *Trends in Cognitive Sciences*, 18(5):235–241, 2014.



- Gerald Jahoda and Judith Schiek Braunagel. *The Librarian and Reference Queries: A Systematic Approach*. Library and Information Science. Academic Press, New York, 1980.
- Maximilian Kalus. Tracing Business Patterns in Sixteenth Century European–Asian Trade: New Methods Using Semantic Networking Models, 2006.
- Maximilian Kalus. Semantic Networks and Historical Knowledge Management: Introducing New Methods of Computer-based Research. *Journal of the Association for History and Computing*, 10 (3), 2007.
- Kris Kiesling. The Influence of American and European Practices on the Evolution of EAD. *Journal of Archival Organization*, 3(2-3):207–215, 2006.
- Jihyun Kim. EAD Encoding and Display: A Content Analysis. *Journal of Archival Organization*, 2 (3):41–55, 2004.
- Michael Krausz. Ideality and Ontology in the Practice of History. In Dussen, W. Jan van der and Lionel Rubinfoff, editors, *Objectivity, Method, and Point of View*, Philosophy of History and Culture, pages 97–111. Brill, Leiden, 1991.
- Carol C. Kuhlthau. Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.
- Rosalie Lack. The Importance of User-Centered Design: Exploring Findings and Methods. *Journal of Archival Organization*, 4(1-2):69–86, 2007.
- Carl Lagoze and Jane Hunter. The ABC Ontology and Model. In Keizo Oyama and Hiro-nobu Gotoda, editors, *Proceedings of the International Conference on Dublin Core and Metadata Applications 2001*, pages 160–176, Tokyo, 2001. National Institute of Informatics.
- William Landis. Archival Outreach on the World Wide Web. *Archival Issues: Journal of the Midwest Archives Conference*, 20(2):129–147, 1995.
- Sebastian Löbner. *Semantik: Eine Einführung*. De-Gruyter-Studienbuch. de Gruyter, Berlin, 2003.
- Chia-Hung Lin, Jen-Shin Hong, and Martin Doerr. Issues in an Inference Platform for Generating Deductive Knowledge: A Case Study in Cultural Heritage Digital Libraries Using the CIDOC CRM. *International Journal on Digital Libraries*, 8(2):115–132, 2008.
- Richard Lytle. Intellectual Access to Archives: I. Provenance and Content Indexing Methods of Subject Retrieval. *The American Archivist*, 43(1):64–75, 1980a.
- Richard Lytle. Intellectual Access to Archives: II. Report of an Experiment Comparing Provenance and Content Indexing Methods of Subject Retrieval. *The American Archivist*, 43(2): 191–207, 1980b.

- Olivia Madison, John Byrum, Jr., Suzanne Jouguelet, Dorothy McGarry, Nancy Williamson, and Maria Witt. Functional Requirements for Bibliographic Records (FRBR): Final Report, 2009.
- William J. Maher. The Use of User Studies. *The Midwestern Archivist*, 11(1):15–26, 1986.
- Kristin E. Martin. Analysis of Remote Reference Correspondence at a Large Academic Manuscripts Collection. *The American Archivist*, 64(1):17–42, 2001.
- Claudio Masolo, Stefano Borgo, Nicola Guarino, Alessandro Oltramari, and Luc Schneider. WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies: Preliminary Report, 2003.
- Willard McCarty. *Humanities Computing*. Palgrave Macmillan, Basingstoke, 2005.
- Dennis Meissner. First Things First: Reengineering Finding Aids for Implementation of EAD. *The American Archivist*, 60(4):372–387, 1997.
- Angelika Menne-Haritz. Access - The Reformulation of an Archival Paradigm. *Archival Science*, 1:57–82, 2001.
- Marvin L. Minsky. *A Framework for Representing Knowledge*, 1974.
- Ani Nenkova and Kathleen McKeown. Automatic Summarization. *Foundations and Trends in Information Retrieval*, 5(2-3):103–233, 2011.
- Cory Nimer and J. Gordon Daines. What Do You Mean It Doesn't Make Sense? Redesigning Finding Aids from the User's Perspective. *Journal of Archival Organization*, 6(4):216–232, 2008.
- Niels Ockeloen, Antske Fokkens, Serge te Braake, Piek Vossen, Victor de Boer, Guus Schreiber, and Susan Legêne. BiographyNet: Managing Provenance at Multiple Levels and from Different Perspectives. In Paul Groth, Marieke van Erp, Tomi Kauppinen, Jun Zhao, Carsten Keßler, Line C. Pouchard, Carole Goble, Yolanda Gil, and Jacco van Ossenbruggen, editors, *Linked Science: In conjunction with the 12th International Semantic Web Conference 2013 (ISWC 2013)*, CEUR Workshop Proceedings, pages 59–71, 2013.
- Dominic Oldman. The CIDOC Conceptual Reference Model (CIDOC-CRM): Primer: Version 1, July 2014, 2014.
- Dominic Oldman, Martin Doerr, Gerald de Jong, Barry Norton, and Thomas Wikman. Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning and Aggregation Services for the Digital Humanities (A Position Paper) System. *D-Lib Magazine*, 20(7/8), 2014.
- Christian-Emil Ore and Øyvind Eide. TEI and Cultural Heritage Ontologies: Exchange of Information? *Literary and Linguistic Computing*, 24(2):161–172, 2009.
- Joy Palmer. Archives 2.0: If We Build It, Will They Come? *Ariadne*, (60), 2009.

- Victoria Peters. Developing Archival Context Standards for Functions in the Higher Education Sector. *Journal of the Society of Archivists*, 26(1):75–85, 2005.
- Daniel V. Pitti. Encoded Archival Description: An Introduction and Overview. *D-Lib Magazine*, 5(11), 1999.
- Daniel V. Pitti and Wendy M. Duff. Introduction. In Daniel V. Pitti and Wendy M. Duff, editors, *Encoded Archival Description on the Internet*, pages 1–7. Haworth Information Press, Binghamton NY, 2001.
- Merrilee Proffitt. How and Why of User Studies: RLG’s RedLightGreen as a Case Study. *Journal of Archival Organization*, 4(1-2):87–110, 2007.
- Christopher J. Prom. The EAD Cookbook: A Survey and Usability Study. *The American Archivist*, 65(2):257–275, 2002a.
- Christopher J. Prom. Does EAD Play Well with Other Metadata Standards? Searching and Retrieving EAD Using the OAI Protocols. *Journal of Archival Organization*, 1(3):51–72, 2002b.
- Mary Jo Pugh. The Illusion of Omniscience: Subject Access and the Reference Archivist. *The American Archivist*, 45(1):33–44, 1982.
- Leopold von Ranke. *Geschichten der romanischen und germanischen Völker: von 1494 bis 1535*, volume 1. Reimer, 1824.
- Clay Redding. Reengineering Finding Aids Revisited: Current Archival Descriptive Practice and Its Effect on EAD Implementation. *Journal of Archival Organization*, 1(3):35–50, 2002.
- Thomas Riechert, Ulf Morgenstern, Sören Auer, Sebastian Tramp, and Michael Martin. Knowledge Engineering for Historians on the Example of the Catalogus Professorum Lipsiensis. In Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, and Birte Glimm, editors, *The Semantic Web*, volume 6497 of *Lecture Notes in Computer Science*, pages 225–240, Berlin, 2010. Springer.
- Andrea Rosenbusch. Are Our Users Being Served? A Report on Online Archival Databases. *Archives & Manuscripts*, 29(1):44–61, 2001.
- James M. Roth. Serving up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids. *The American Archivist*, 64(2):214–237, 2001.
- Jörn Rüsen and Friedrich Jaeger. Historische Methode. In Richard van Dülmen, editor, *Geschichte*, Das Fischer Lexikon, pages 13–32. Fischer Taschenbuch Verlag, Frankfurt am Main, 1990.
- Tuukka Ruotsalo, Lora Aroyo, and Guus Schreiber. Knowledge-Based Linguistic Annotation of Digital Cultural Heritage Collections. *IEEE Intelligent Systems*, 24(2):64–75, 2009.

- Alistair Russell, Paul R. Smart, Dave Braines, and Nigel R. Shadbolt. NITELIGHT: A Graphical Tool for Semantic Query Construction, 2008.
- Gerard Salton and Michael J. MacGill. *Information Retrieval: Grundlegendes für Informationswissenschaftler*. McGraw-Hill-Texte. McGraw-Hill, Hamburg, 1987.
- Jennifer Schaffner. *The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies*. OCLC Programs and Research, Dublin, Ohio, 2009.
- Robert Jones Shafer and David Harry Bennett, editors. *A Guide to Historical Method*. The Dorsey series in history. Dorsey Press, Homewood, 3 edition, 1980.
- Ryan Shaw. Information Organization and the Philosophy of History. *Journal of the American Society for Information Science and Technology*, 64(6):1092–1103, 2013.
- Donghee Sinn. Room for Archives? Use of Archival Materials in No Gun Ri Research. *Archival Science*, 10(2):117–140, 2010.
- Richard P. Smiraglia. Subject Access to Archival Materials Using LCSH. *Cataloging & Classification Quarterly*, 11(3-4):63–90, 1990.
- John F. Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks Cole, Pacific Grove, 2000.
- Susan Leigh Star. The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In Les Gasser and Michael N. Huhns, editors, *Distributed Artificial Intelligence*, Research notes in artificial intelligence, pages 37–54. Pitman, London, 1989.
- Robert Stevens, Alan Rector, and Duncan Hull. What is an Ontology?, 2010. URL <http://ontogenesis.knowledgeblog.org/66>.
- Richard V. Szary. Encoded Finding Aids as a Transforming Technology in Archival Reference Service. In Daniel V. Pitti and Wendy M. Duff, editors, *Encoded Archival Description on the Internet*, pages 187–197. Haworth Information Press, Binghamton NY, 2001.
- Robert S. Taylor. The Process of Asking Questions. *American Documentation*, 13(4):391–396, 1962.
- Kate Theimer. What is the Meaning of Archives 2.0? *The American Archivist*, 74(Spring/Summer): 58–68, 2011.
- Helen R. Tibbo. *Abstracting, Information Retrieval, and the Humanities: Providing Access to Historical Literature*, volume 48 of *ACRL Publications in Librarianship*. American Library Association, Chicago, 1993.

- Helen R. Tibbo. Primarily History in America: How U.S. Historians Search for Primary Materials at the Dawn of the Digital Age. *The American Archivist*, 66(1):9–50, 2003.
- Helen R. Tibbo and Gary Marchionini. Primarily History: Historians and the Search for Primary Source Materials. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2002*, pages 1–10, New York, NY, 2002. ACM.
- John Tosh. *The Pursuit of History: Aims, Methods, and New Directions in the Study of Modern History*. Longman, London, 2nd ed., 4. impr edition, 1996.
- Ciaran B. Trace and Andrew Dillon. The Evolution of the Finding Aid in the United States: From Physical to Digital Document Genre. *Archival Science*, 12(4):501–519, 2012.
- Katerina Tzompanaki and Martin Doerr. A New Framework for Querying Semantic Networks: Museum and the Web 2012, April 11-14, 2012, San Diego, CA, USA, 2012a.
- Katerina Tzompanaki and Martin Doerr. Fundamental Categories and Relationships for Intuitive Querying CIDOC-CRM Based Repositories: Technical Report ICS-FORTH/TR-429, 2012b.
- Karl J. Weintraub. The Humanistic Scholar and the Library. *Library Quarterly*, 50(1):22–39, 1980.
- Thomas D. Wilson. Human Information Behavior. *Informing Science*, 3(1):49–55, 2000.
- Thomas D. Wilson. On User Studies and Information Needs. *Journal of Documentation*, 62(6): 658–670, 2006.
- Wilhelm Windelband. *Geschichte und Naturwissenschaft*, volume 1894 of *Rektoratsreden der Universität Strassburg*. Heitz, Strassburg, 3. unveränd. aufl. edition, 1904.
- Sonia Yaco. It's Complicated: Barriers to EAD Implementation. *The American Archivist*, 71 (Fall/Winter):456–475, 2008.
- Elizabeth Yakel. Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users? *Journal of Archival Organization*, 2(1-2):63–77, 2004.
- Gideon Zenz, Xuan Zhou, Enrico Minack, Wolf Siberski, and Wolfgang Nejdl. From Keywords to Semantic Queries: Incremental Query Construction on the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):166–176, 2009.
- Junte Zhang and Jaap Kamps. Search Log Analysis of User Stereotypes, Information Seeking Behavior, and Contextual Evaluation. In Nicholas J. Belkin and Diane Kelly, editors, *Proceedings of the Third Symposium on Information Interaction in Context*, pages 245–254, New York, 2010. ACM.

Junte Zhang, Jaap Kamps, Gottfried Vossen, Darrel D. E. Long, and Jeffrey Xu Yu. Focused Search in Digital Archives. In Gottfried Vossen, Darrell D. E. Long, and Jeffrey Xu Yu, editors, *Web Information Systems Engineering - WISE 2009*, Lecture Notes in Computer Science, pages 463–471, Berlin, 2009. Springer.

Xiaomu Zhou. Examining Search Functions of EAD Finding Aids Web Sites. *Journal of Archival Organization*, 4(3-4):99–118, 2007.

---

*All URLs in this study have been checked and updated on July 30th 2017.*

## **Danksagung**

Größter Dank gebührt den zahlreichen über die Jahre Involvierten und Tangierten für großzügige und vielfältige Unterstützung sowie stets geduldige Nachsicht! Im Besonderen gilt mein Dank Stefan Gradmann, der mich ermutigt und erst die Möglichkeit eröffnet hat, eine informationswissenschaftliche Dissertation zu schreiben. Meinen beiden Betreuern, Vivien Petras und Martin Doerr, möchte ich ausdrücklich für ihre wertvolle Unterstützung danken, insbesondere auch Martin Doerr für die langen und intensiven Gespräche. Angelika und Klaus, Pam, Janna und Darko haben mit ihrem unerschütterlichen Zuspruch diese Arbeit erst ermöglicht. Nicht zu vergessen ist schließlich die aufopferungsvolle Arbeit der zahlreichen Korrekturleser, besonders die von Gerold.