Zellīj:
# A semantic pattern development and documentation system

George Bruseker (Takin.solutions)
CRM SIG, Luxembourg, 7/12/2022

What are we doing?

Semantic Data patterns database and documentation workflow system - a way of organizing semantic data patterns and sharing them, enabling semantic data to be generated and reused in a practical and sustainable way.

You get:
- **the expressive flexibility of semantics**
- **the structural reliability of traditional data standards**
- **a way to consistently and effectively document all that**
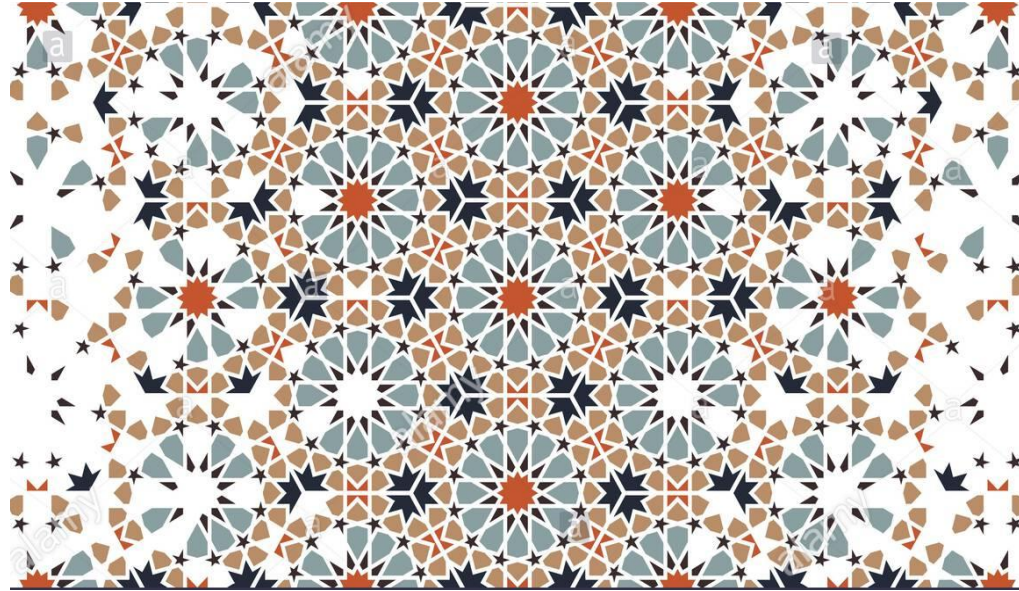
Background: SRDM
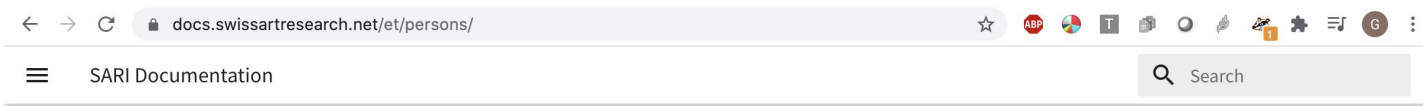
Person Reference Data Model

Author: George Bruseker, Nicola Carboni

Version: 1.0

Introduction

The person reference data model provides a list of standard descriptors (fields) that are typically present in the description of a person in cultural heritage data systems. The intention of this reference model is to provide a consolidated, high-level formal structure comprising the most commonly reused descriptors for a person entity and, further, to provide for these a set of semantic mapping to the CIDOC CRM. Moreover, each field is marked with respect to its potential functionality with regards to instance matching between overlapping datasets. This reference data model aims to serve a number of functions including:

1. to support the creation of semantic data entry forms that will create 'born semantic' data streams

2. to serve as a reference guide in designing eventually compatible data structures for feeding into larger aggregations of Cultural Heritage data

3. to guide mapping processes of extant data sources with common mapping patterns

4. to be adopted by institutions acting as aggregation hubs in order to create consistent re-expressions of extant reference information in a common CIDOC CRM based expression

SRDM: Aims / Successes

- Gives a simple presentation of reference object in form understandable to domain expert but creating semantics for robustness of data
- Creates a guide for consistent application of semantics
    - To create new data OR
    - Map old data
- Continue to manage your data in its source form
- Solves the simple semantic problems consistently, so you can focus on the tricky ones

Ok, great, but...

# New Challenges for Semantic Documentation

The success of the idea of Semantic Reference Data Models opens new challenges:

- Extension of models
- Revision/versioning of models
- Efficiency and Accuracy of Operation
- Full usability by all user types
- Adoption and proposals by other projects and broader community
- Attribution of authorship/provenance/funding support
- Authority

# So… Zellīj

Zellīj: what, more specifically, is it?
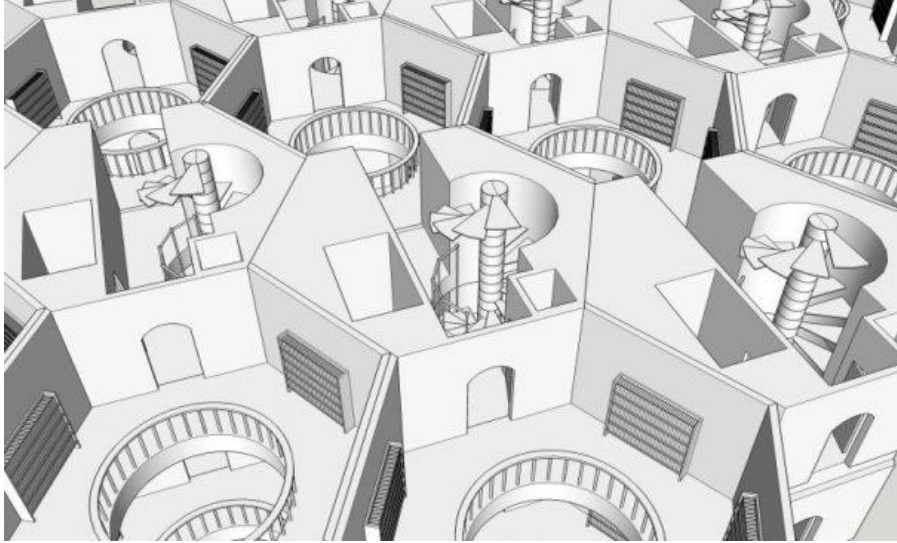
A database and workflow that allows the documentation of semantic patterns:

- At different levels of granularity
- In a compositional manner
- With researcher appropriate labels and descriptions
- With developer appropriate description of the semantic pattern rdf notation
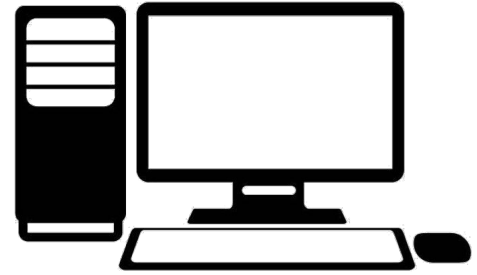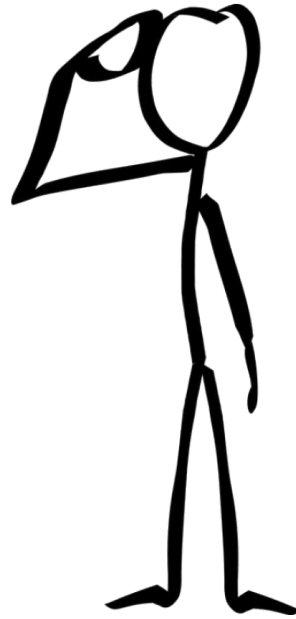- With attribution of creation, funding, use

Who is it for?
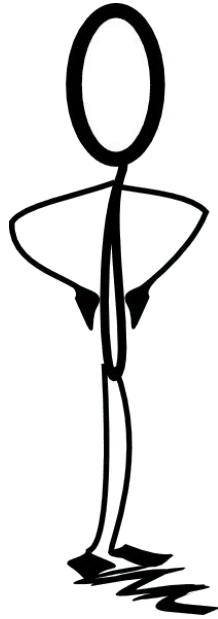
Domain Specialist   Software Developer /   Conceptual Modeller   Machines
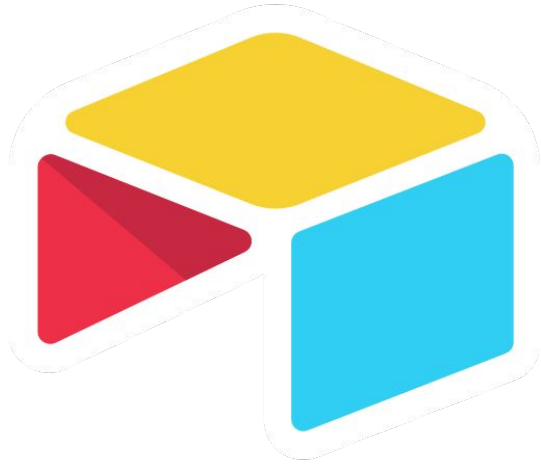Systems Implementer

Why are we doing it?

Building:

- Share Semantics Knowledge
- Increase Efficiency of Semantics Use
- Build Semantics Community
- Get benefits of Semantics without the Problems
- Create Sustainability

Towards:

- Share Research Data / Support New Knowledge Creation

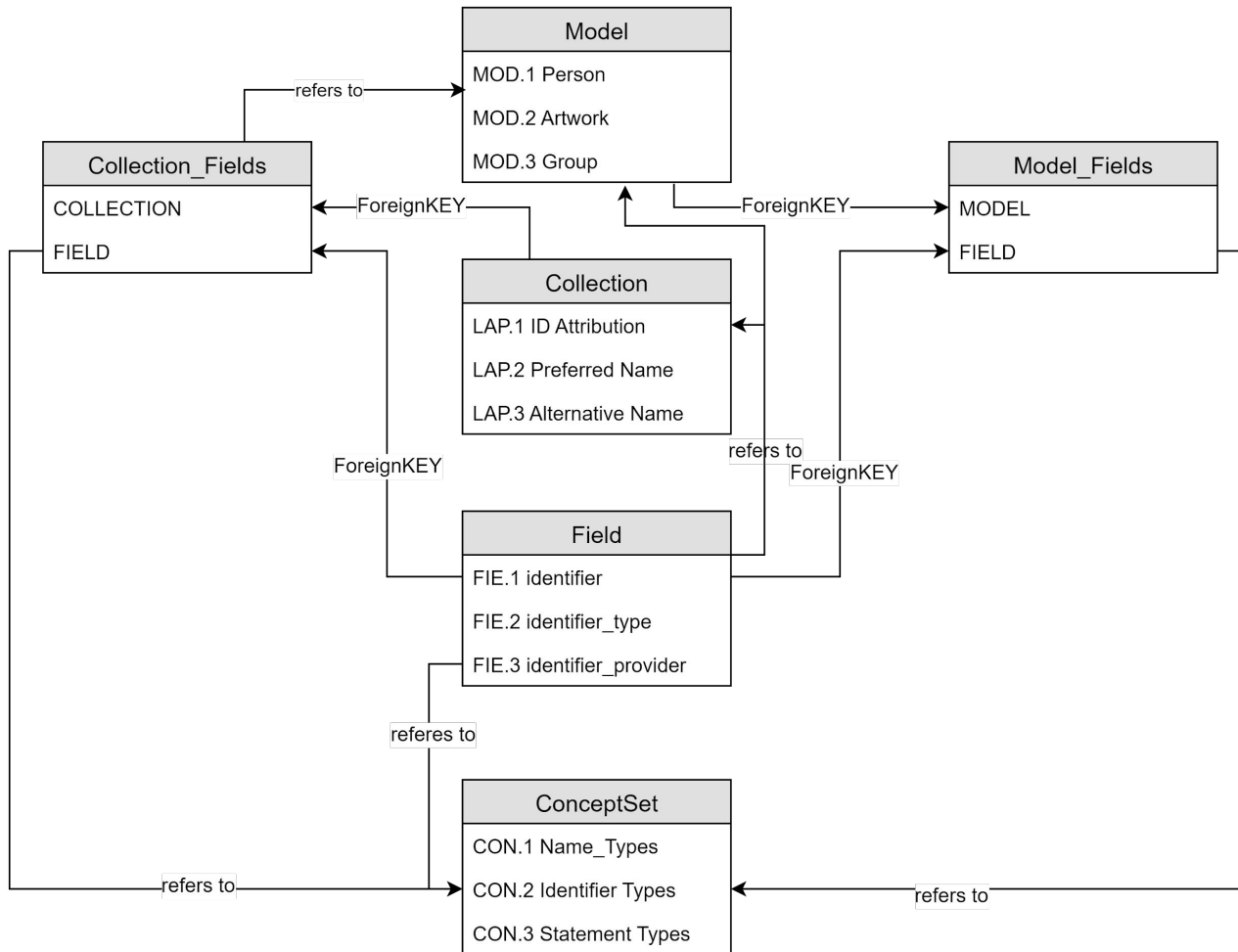# System Software

Backend

Frontend

What does database consist of?

Fields: a named, ided, and intensionally defined, syntactically unique path from a domain node to a range node through *n* triples, representing a semantically unique statement type, intended to represent one value / data point

Collections: a named, ided, and intensionally defined grouping of fields composing a pattern of mutually relevant statement types related to the same, intermediate subject (domain node), documented for the purpose of common reuse across entities in a semantic data project

**Models**: a named, ided and intensionally defined grouping of fields composing a pattern of mutually relevant statement types related to the same, central subject (domain node), documented for the purpose of common reuse across semantic data projects

Category: a named cluster of fields within a model grouped together according to their likelihood/capacity to respond to a similar field of inquiry

Field

Identifier	Field ID	Name	Description	Ontological Scope	CRM Path	Total_Turtle	Expected Value Type	Expected Resource Model	Expected Collection Model	Concept - Specified	Version
FIE.1	00dba278	Identifier	This field is used to record an identifier attributed to the documented entity.	E1 CRM Entity	--> P1 --> E42[1_1] --> P190 --> rdf:literal	<http://www.srdm.org/object/E1> a <http://www.cidoc-crm.org/cidoc-crm/E1_CRM_Entity>; crm:P1_is_identified_by <https://www.srdm.org/name/fie_1_1> . <https://www.srdm.org/name/fie_1_1> a crm:E42_Identifier; crm:P190_has_symbolic_content "content" .	String				1
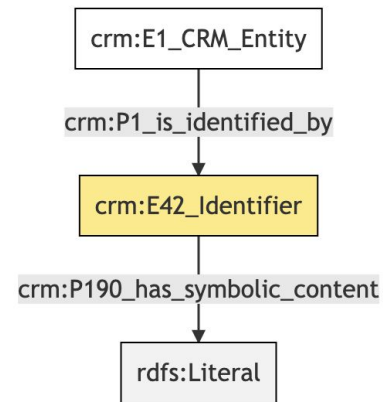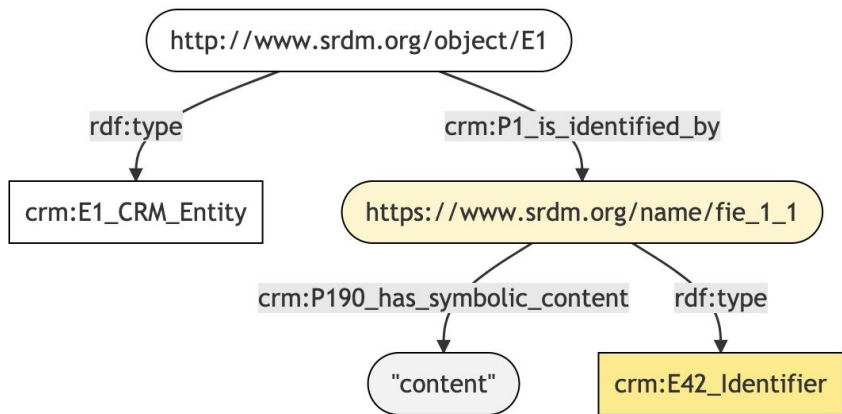
Collection

Identifier	uuid	Name	Description	Ontological Scope	Collection_Fields	Authors	Funders	Users
COL.1	1ac1aa	ID Attribution	A series of attributes used to indicate an identifier associated to a document object, its attributes and the provenance thereof.	E1 CRM Entity	[COL.1] ID Attribution_[FIE.1] Identifier,[COL.1] ID Attribution_[FIE.2] Identifier Type,[COL.1] ID Attribution_[FIE.3] Identifier Provider,[COL.1] ID Attribution_[FIE.4] Identifier Source	George Bruseker,Nicola Carboni	Swiss Art Research Institute	Swiss Art Research Institute
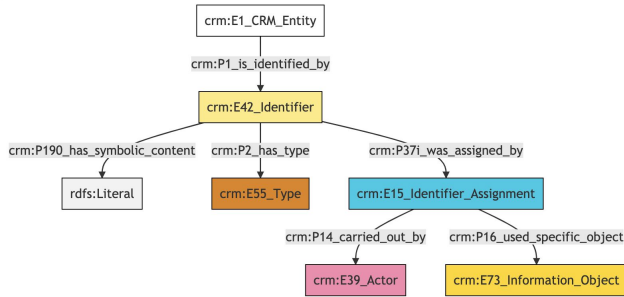

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .

<http://www.srdm.org/object/E1>
  a <http://www.cidoc-crm.org/cidoc-crm/E1_CRM_Entity> ;
  crm:P1_is_identified_by <https://www.srdm.org/name/fie_1_1> .

<https://www.srdm.org/name/fie_1_1>
  a crm:E42_Identifier ;
  crm:P190_has_symbolic_content "content" ;
  crm:P2_has_type <http://www.srdm.org/type/fie_2_1> ;
  crm:P37i_was_assigned_by <https://www.srdm.org/event/fie_3_1> .

<http://www.srdm.org/type/fie_2_1> a ns0:E55_Type .
<https://www.srdm.org/event/fie_3_1>
  a crm:E15_Identifier_Assignment ;
  crm:P14_carried_out_by <https://www.srdm.org/actor/fie_3_2> ;
  crm:P16_used_specific_object <https://www.srdm.org/conceptual_object/fie_4_1> .

<https://www.srdm.org/actor/fie_3_2> a ns0:E39_Actor .
<https://www.srdm.org/conceptual_object/fie_4_1> a ns0:E73_Information_Object .
```
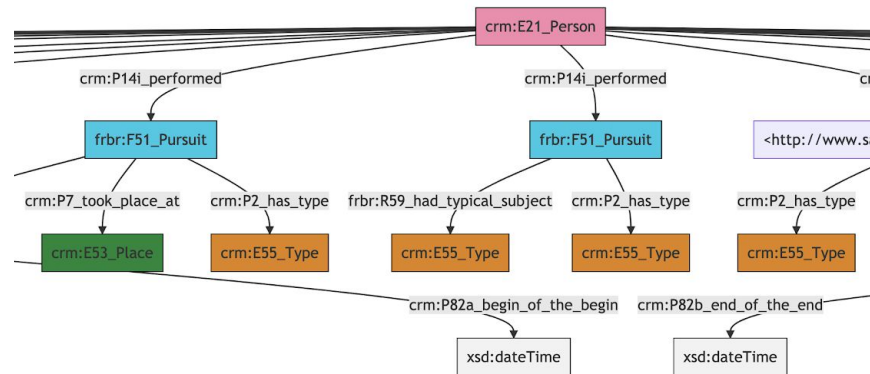
Model

Identifier	Name	Description	Ontological Scope	Model_Fields	Authors	Funders	Users	Version	Status
MOD.1	Person	The person reference data model provides a list of standard descriptors (fields) that are typically present in the description of a person in cultural heritage data systems. The intention of this reference model is to provide a consolidated, high-level formal structure comprising the most commonly reused descriptors for a person entity and, further, to provide for these a set of semantic mapping to the CIDOC CRM. Moreover, each field is marked with respect to its potential functionality with regards to instance matching between overlapping datasets.	E21 Person	[MOD.1] Person_[FIE.1] Identifier,[MOD.1] Person_[FIE.2] Identifier Type,[MOD.1] Person_[FIE.3] Identifier Provider,[MOD.1] Person_[FIE.4] Identifier Source,[MOD.1] Person_[FIE.5] Name,[MOD.1] Person_[FIE.6] Name Language,[MOD.1] Person_[FIE.7] Name Part,[MOD.1] Person_[FIE.8] Name Part Type,[MOD.1] Person_[FIE.10] Alternative Name,[MOD.1] Person_[FIE.11] Alternative Name Type,[MOD.1] Person_[FIE.12] Alternative Name Language,[MOD.1] Person_[FIE.13] Alternative Name Use Period - ...	George Bruseker,Nicola Carboni	Swiss Art Research Infrastructure	Swiss Art Research Infrastructure	1.0	Released


```
...
<http://www.srdm.org/actor/E21>
  a <http://www.cidoc-crm.org/cidoc-crm/E21_Person> ;
  crm:P1_is_identified_by <https://www.srdm.org/name/fie_1_1>,
<https://www.srdm.org/name/fie_5_1>,
<https://www.srdm.org/name/fie_10_1> ;
  crm:P98i_was_born <https://www.srdm.org/event/fie_73_1> ;
  crm:P100i_died_in <https://www.srdm.org/event/fie_76_1> ;
  crm:P107i_is_current_or_former_member_of
<https://www.srdm.org/actor/fie_145_1>
...
```

Automated Documentation

Description

Identifier	Name	Function / Description	CRM Path
['fie_188']	['Group Description']	['This field is used to record a description in free text of the documented group.']	['--> P129i --> E33[188_1] --> P190 --> rdfs:Literal']
['fie_190']	['Group Description Type']	['This field is used to record the type of description given of the documented group.']	['--> P129i --> E33[188_1] --> P2 --> E55[190_1]']
['fie_191']	['Group Description Author']	['This field is used to record the author of the description given of the documented group.']	['--> P129i --> E33[188_1] --> P94i --> E65[191_1]>-P14 --> E39[191_2]']
['fie_192']	['Group Description Earliest Date']	['This field is used to record the earliest possible date for the creation of the description of the documented group.']	['--> P129i --> E33[188_1] --> P94i --> E65[191_1] --> P4 --> E52[191_2] --> P82a --> xsd:dateTime']
['fie_193']	['Group Description Latest Date']	['This field is used to record the latest possible date for the creation of the description of the documented group.']	['--> P129i --> E33[188_1] --> P94i --> E65[191_1] --> P4 --> E52[191_2] --> P82b --> xsd:dateTime']

⌐ Ontology Graph **⊹ Instance Graph** **</> Turtle RDF** **≡ JSON-LD**

Generated by Zellij, © 2021.
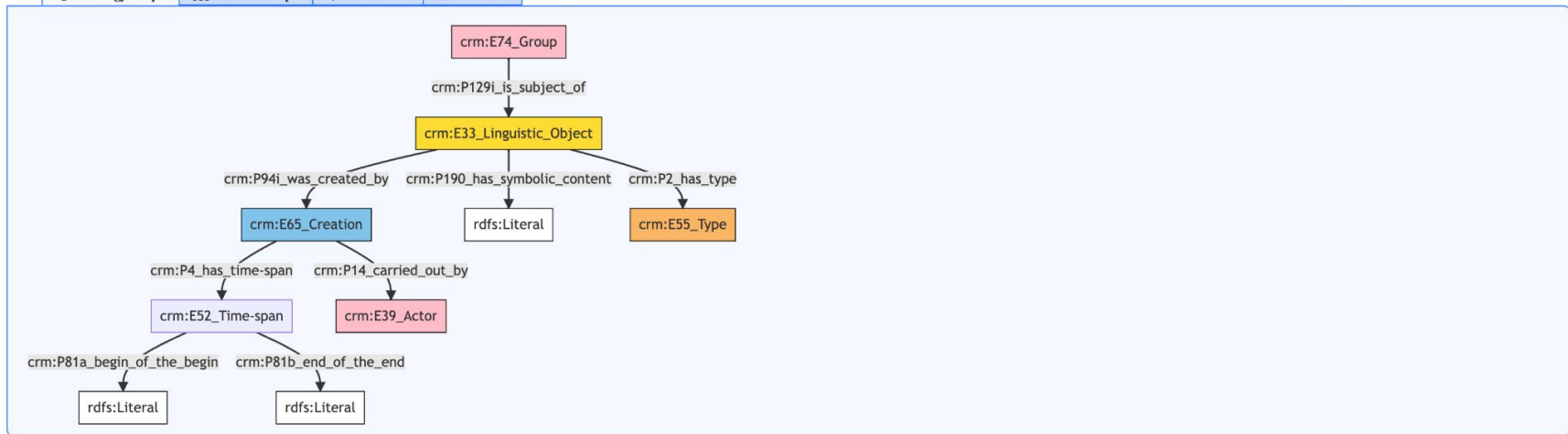
SARI USE Case

Fields: 202 distinct base patterns

Collections: 43 distinct grouping patterns

Models: 8 fully documented standard semantic reference data models

Fields Breakdown

Ontological Scope	Count	Grouping
E21 Person	21	Actor
E74 Group	15	Actor
E39 Actor	17	Actor
D1 Digital Object	13	Conceptual Object
F24 Publication Expression	11	Conceptual Object
E89 Propositional Object	3	Conceptual Object
E2 Temporal Entity	11	Event
E7 Activity	9	Event
E4 Period	3	Event
E1 CRM Entity	38	General
E72 Legal Object	3	General
E18 Physical Thing	20	Physical Thing
E24 Physical Man-Made Thing	24	Physical Thing
E53 Place	14	Place

Count

- Place 6.9%
- Physical Thing 21.8%
- Actor 26.2%
- Conceptual Object 13.4%
- Event 11.4%
- General 20.3%
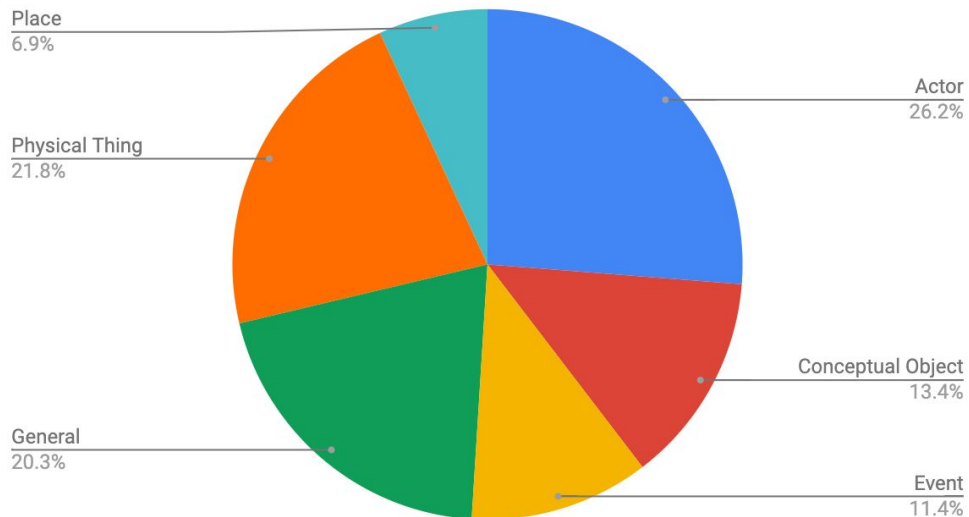
Collections Breakdown

Ontological Scope	Count	Grouping
E21 Person	4	Actor
E74 Group	3	Actor
E39 Actor	4	Actor
D1 Digital Object	3	Conceptual Object
F24 Publication Expression	2	Conceptual Object
E2 Temporal Entity	4	Event
E7 Activity	1	Event
E4 Period	1	Event
E1 CRM Entity	6	General
E72 Legal Object	1	General
E24 Physical Man-Made Thing	5	Physical Thing
E18 Physical Thing	3	Physical Thing
E53 Place	2	Place

Count

Place 5.1%
Physical Thing 20.5%
General 17.9%
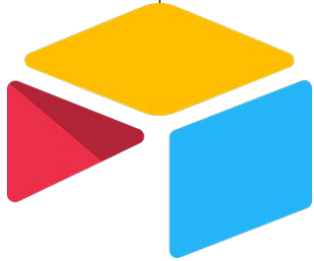Actor 28.2%
Conceptual Object 12.8%
Event 15.4%
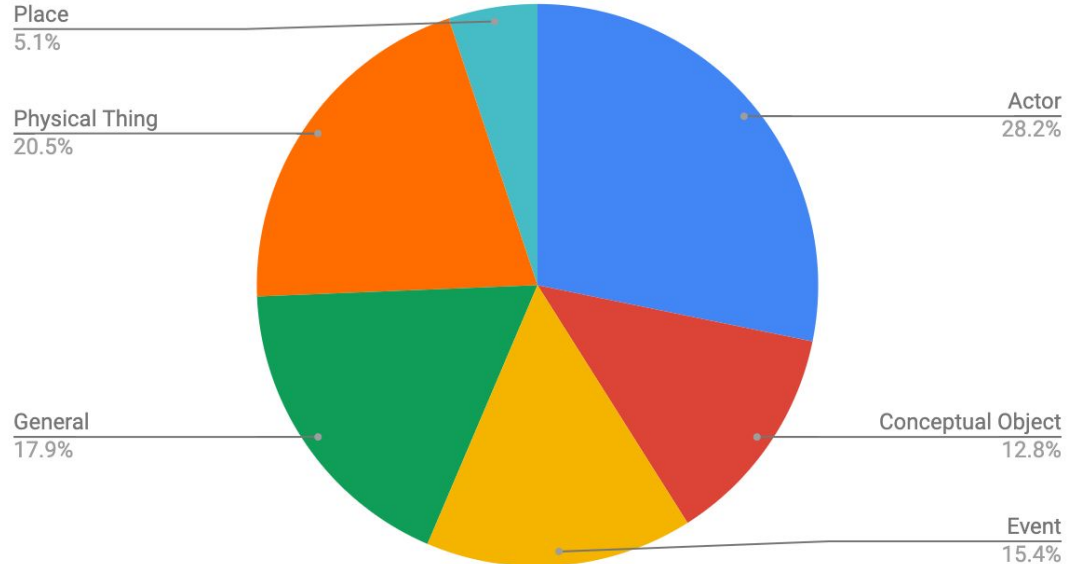
High level
Documentation
Architecture
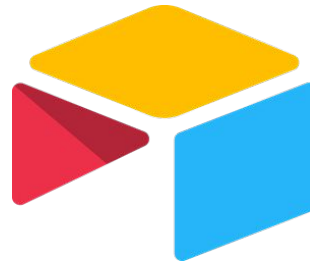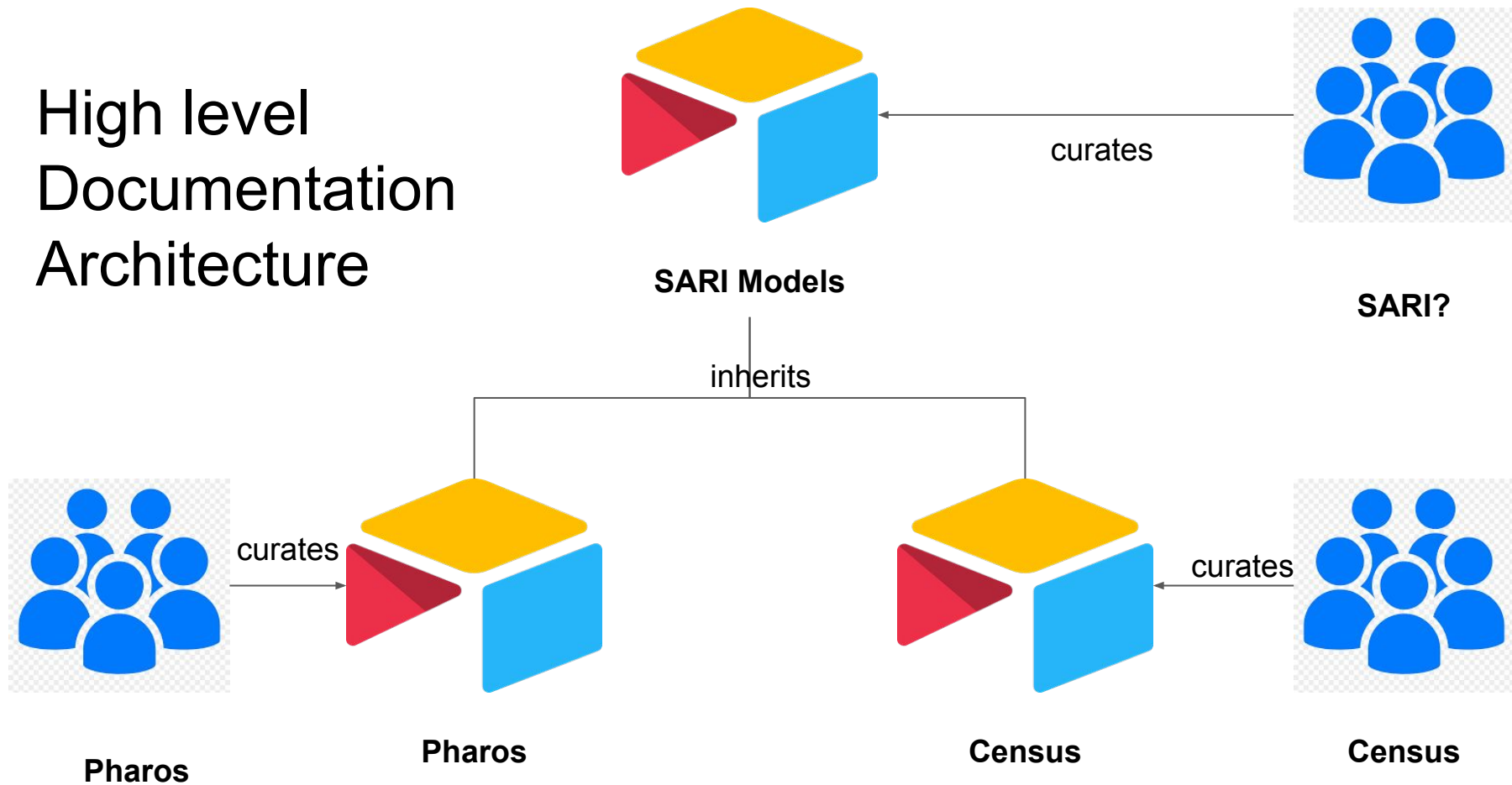
SARI

inherits

Pharos
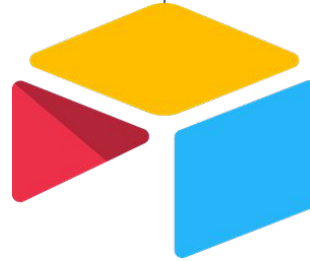
Hertziana
Semantic

Semantic
Census

What have we achieved?

- A data model for a reusable database of patterns
- A database of basic, compositional patterns (fields, collections)
- A database with a set of standard models built compositionally (8)
- A workflow for generating consistent documentation

What is there still to do?

- Make whole platform open source
- Extend with Canned Sparql
- Extend with SHACL
- Extend with auto build out of system templates for Arches / Research Space / Wisski etc.
- Connect to Ontology production systems like OntoMe

THANK YOU