# Statements about Statements.

CRM SIG community draft, May 2023.

Current state: Some preliminary thoughts for initialising proper elaboration.

## Problem

When following the evolution of a discourse about some “universe”, in our case “historical sciences” that aim at shedding light what has or may have happened in the past, how and why, we are confronted with the problem that each “justified belief” achieved by someone becomes *a historical fact* in its own right. This “*provenance of knowledge” is mandatory* for any scientific progress, regardless whether in humanities or sciences in the narrower sense.

In terms of logic and within the domain of knowledge representation by KR languages and formal ontologies, we can technically simplify the question to how to make statements about statements as part of a knowledge graph. Unfortunately, “classical” logic has not dealt with the problem. There is a significant *lack of theoretical* and philosophical support.

However, given the more and more pressing practical needs gave rise to a set of *existing technical* constructs and a vivid discussion about pros and cons, and various communities defending their favourite constructs. The title of a paper “Don’t Like RDF Reification? Making Statements about Statements Using Singleton Property” reveals the emotional flavour of the discussion. A major encoding problem is the fact that the identifier of property instances cannot be referred to directly in most current KR languages.

A good technical survey is:

Sikos, Leslie F., and Dean Philp, ‘Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs’, Data Science and Engineering, 5.3 (2020), 293–316 <https://doi.org/10.1007/s41019-020-00118-0>

And “Representing provenance and track changes of cultural heritage metadata in RDF: a survey of existing approaches” https://arxiv.org/abs/2305.08477

Since CRM-SIG is committed to recommendations that give priority to the ontological-epistemological question and tries to be neutral to technological “fashions”, in this issue we try to analyse the problem from the following sides:

1. The ontological-epistemological requirements in historical sciences
2. Possible formal representations, even when a conclusive theory is lacking
3. Candidate implementations using technology at hand – expressive effectiveness
4. Benchmarking scalability and efficiency for data entry, storage, querying and result representation.

As characteristic examples, let us take the following:

E1) “as designer”

E2) Two historical opinions about Nero’s whereabouts, and the related discourse of historical research:

Nero in July 19, 64 AD (E93 Presence)

 P164 is temporally specified by: July 19, 64 AD (E52 Timespan)

 P195 was a presence of: Nero Claudius Caesar Drusus Germanicus (E21 Person)

 P167 was within Antium, Italy (E53 Place)

 P133 is spatiotemporally separated from: The Great Fire of Rome (E5 Event)

P1 is indentified by: incendium magnum Romae (E41 Appellation)

P4 has timespan: July 19-27, 64 AD (E52 Timespan)

P7 took place at : Rome in 64AD, Italy (E53 Place)

Nero July 19, 64 AD (E93 Presence)

 P164 is temporally specified by: July 19, 64 AD (E52 Timespan)

 P195 was a presence of: Nero Claudius Caesar Drusus Germanicus (E21 Person)

 P167 was within Rome in 64AD, Italy (E53 Place)

P10 falls within (contains): Nero Singing (E7 Activity)

 P2 has type: Singing (E55 Type)

 P14 carried out by: Nero Claudius Caesar Drusus Germanicus (E21 Person)

 P4 has timespan: July 19, 64 AD (E52 Timespan)

 P7 took place at : Rome in 64AD, Italy (E53 Place)

 P132 spatiotemporally overlaps with: The Great Fire of Rome (E5 Event)

P1 is indentified by: incendium magnum Romae (E41 Appellation)

P4 has timespan: July 19-27, 64 AD (E52 Timespan)

P7 took place at : Rome in 64AD, Italy (E53 Place)

E3) an integration scenario, in which one resource wants to update a contribution which partially contains identical property instances with another resource.

# ontological-epistemological requirements

**The ontological problem** may be described as the duality between reference and its use about referents, or between information and knowledge, or between uttering and believing, well known from semiotics.

In more detail: A particular text must be seen as an object of historical reality, created at a specific time and place by specific people. It has a symbolic form and form-based identity, and may be found or have been found on one or more material carriers. At the same time, it may consist of statements, also called “references”, about some universe of discourse, its meaning, also called “referents”, such as parts of historical reality (or, less relevant for us, mathematics or fictional plots). Following Wittgenstein’s Tractatus, such statements can be related to “states of affairs” in the respective reality. For instance, (and in particular), formal propositions in a knowledge base are understood as claims about relations between things users can identify in reality, such as the Louvre, Mona Lisa the painting, or things we reasonably assume to having existed, such as Emperor Qin Shi Huang of China, the library of Alexandria. Each such particular relation can be regarded to have an identity as a meaning.

Vice versa, all propositions, formal or not, about some reality (“referents”) have to be communicated in some symbolic form (encoded “references”), which thereby have a second, form-based identity. The meaning of context-free propositions, such as “Martin Doerr – has worked at – FORTH in Heraklion” can be regarded having one identity independent from the utterance, a composite of the identities of the referred entity instances (Martin Doerr, FORTH in Heraklion) and the property type (has worked at), with the silent assumption that users can relate these in a compatible way to their reality. Propositions with the same meaning may appear many times as an utterance, with different authorship and history, and in varying, equivalent symbolic forms.

Based on this theory, RDF-OWL compatible knowledge bases conflate all equivalent propositions into one internal representation that appears in queries. XML databases do exactly the opposite: they preserve only the unit (“document”) given by authorship, but allow for querying equivalent meanings appearing in multiple documents. “Quad stores”, on the other side, maintain internally a link between each equivalence set of propositions and the contributing “contexts”, typically the documents by which RDF-OWL statements have been loaded.

This brings us to the last ontological question: Can the intended meaning of statements in a document created in a specific context be reduced to consider contained “atomic propositions”, one by one; or is the whole of propositions in a document “more than its parts”? To our best knowledge, wholes in general are more than its parts, and there exist serious philosophical arguments that considerations of truth (or approximate truth, degrees of truth etc.) of knowledge cannot be reduced to single propositions considered in isolation. Therefore we maintain that the sets of propositions which have historically been made together constitute an ontological entity in their own right.

Summarizing:

1. There is a **one-to-many** relationship between the (identity of the) meaning of a formal proposition and the symbolic form in which it may have been stated in different contexts.
2. The meaning of a formal proposition can *only be communicated* and documented by an equivalent symbolic form.
3. The context at utterance time should determine whether a real-world meaning *was intended to* be communicated by the proposition (see also Wittgenstein Tractatus). The *context of use* determines when the symbolic form should be treated as the intended meaning or as an information object.
4. **Not only** the identity of meaning of a single proposition must be considered, but also *the sets of propositions* historically been made together for some purpose.

**The epistemological problem** can be described as the processes of knowledge acquisition, reasoning and documentation in which propositions and proposition sets need to be referred to or used both by meaning and by context of utterance as described above. Constrained to the maintenance of knowledge bases, these considerations should be the base for recommending adequate kinds of technical solutions, current or expected, depending on the use case.

We may distinguish:

1. Specific reasoning and justified opinions about individual facts, as appearing during research, being targeted to the subject or by chance discovering issues that deserve settling.
2. Observations, in systematic field studies, in laboratories, as witness of historical events etc.
3. A self-contained, internally consistent study or report by a team of authors finished or “fixed” at a particular time, such as scientific articles, an exhibition catalogue, the documentation of a particular museum object by one curator, etc.
4. Citations and references to other work, being believed or questioned.
5. Units of data entry in a knowledge base or other encyclopaedic framework.
6. Integrating information from multiple databases.

We should consider the units (single property – proposition set) pertaining to a single, identifiable context of knowledge creation, transfer, reasoning and the scope of respective belief values.

# Possible formal representations

We may consider:

E13 Attribute Assignment with single target property,

Propositions Sets as “interpreted information objects”.

Annotation models specialized to propositions as target…

E13 Attribute Assignment extended to multiple properties

 (such as:

Ixx1 Proposition Set Assignment

 subclassOf: E7 Activity

 superclassOf: E13 Attribute Assignment

Jxx2 assigned propositions (propositions assigned by)

 domain: Ixx1 Proposition Set Assignment

 range: I4 Proposition Set

Ixx2 Single Property

 subclassOf: I4 Proposition Set

Jxx4 has domain (is domain of)

 domain: Ixx2 Single Property

 range: E1 CRM Entity

Jxx5 has range(is range of)

 domain: Ixx2 Single Property

 range: E1 CRM Entity

Jxx6 has type(is type of)

 domain: Ixx2 Single Property

 range: E55 Type

P140 assigned attribute to (was attributed by)

 P140(x,y) <=> Jxx2 (x,z) AND Ixx2(z) AND Jxx4(z,y)

P141 assigned (was assigned by)

 P141(x,y) <=> Jxx2 (x,z) AND Ixx2(z) AND Jxx5(z,y)

 P177 assigned property of type (is type of property assigned)

 P177(x,y) <=> Jxx2 (x,z) AND Ixx2(z) AND Jxx6(z,y)

…..)

# Candidate implementations

We may consider implementations such as:

Such as RDF reification running on XXX, … Serialization formats, Named Graphs as “contexts” on Virtuoso etc., RDF\* …

# Benchmarking

scalability and efficiency for data entry, storage, querying and result representation.