A New Framework for Querying Semantic Networks

Katerina Tzompanaki[1], Martin Doerr[1]

[1] Institute of Computer Science, F.O.R.T.H.
N. Plastira 100 Vassilika Vouton
GR-700 13 Heraklion, Crete, Greece
{katetzob, martin}@ics.forth.gr

TECHNICAL REPORT: ICS-FORTH/TR-419, May 2011

Abstract. Open World information systems, such as Digital Libraries or the Web, can successfully be accessed by text search engines. Semantic networks of RDF triples however, the core of the Semantic Web, are not easily accessible by such a global query paradigm. Querying individually hundreds of different kinds of properties leaves a huge recall gap to text retrieval, whereas a global restriction to "core metadata" deprives the systems of reasoning capability. We propose and are implementing a new query paradigm: Generic relationships, as we are used to from core metadata, are presented to the user in the form of a system of complex deductions from a rich underlying network of more specialized actual metadata, rather than being primary documentation elements. Using as schema the core ontology ISO21127 and specializations of it, we are able to achieve high recall with a compact set of very simple, comprehensive questions in an application comprising general cultural heritage data and digital provenance data of specific 3D modeling processes. Application of the framework can easily be adjusted to many domains and user preferences.

Keywords: Semantic networks, information access, semantic search, metadata, reasoning

1 Introduction

Till recently, query languages, such as SQL, and query user interfaces were developed for 'Closed World' systems, such as accounting systems, where information is comprehensively described within the limits of a context and a schema well-known to the users. In such systems, querying by associations of values in different database fields yields very high recall and precision. In 'Open World' system, such as Digital Repositories or the Web, information may be organized by different people, using a schema in different ways, or even using different schemata and languages, and information is by nature incomplete. Therefore the traditional query paradigm becomes much more unreliable, which may explain why "advanced search" options are so unpopular.

Currently, the most popular search method in Open World systems is the keyword based search in text documents, image captions and database fields. It usually yields high recall rate and a medium to low precision rate. The optimization of recall **and** precision is subject of intensive information retrieval research. The big search engines on the Internet may find millions of "hits" for some term, but only few documents may actually be the ones sought. Query term expansion using a thesaurus of synonyms and related terms may improve the recall, but may further deteriorate the precision. The user satisfaction is nevertheless relatively high, because the system is very "responsive" to the users' requests, in particular, if a smart relevance ranking system is used to increase precision. But the satisfaction is mostly due to the huge number of redundant data in these systems with respect to the most popular questions.

The Semantic Web promises to overcome the recall and precision problem for information not backed up by high redundancy, by resorting to rich formally structured metadata for documents of interest and links between them, and combining them with general formal background knowledge. The data are formulated in the "Semantic Web Languages" RDF and OWL under schemata ("ontologies") that are globally accessible via Internet and can be combined to a certain degree. Implementations create "semantic networks" on database managers called RDF Triple Stores. The most advanced Digital Library systems, such as the Europeana[1] or cultureSampo[2], are based on this technology. However, the Semantic Web is an Open World system. Querying individually hundreds of different kinds of properties creates a huge recall gap compared to text retrieval, and querying a conjunction of even a few properties uses to frustrate the users with empty answers (see also [27]). A global restriction of the semantic network to "core metadata" on the other side deprives the systems of the reasoning capability and precision the Semantic Web promised.

In order to fill this gap of precision and recall rates between keyword search and semantic search on metadata, we propose a querying system for semantic networks based on a few "Fundamental Categories and (binary) Relationships". These categories ("FCs" in the following) are base classes covering the domain and the relationships ("FRs") are deductions from complex path expressions in a much richer and more specialized semantic network, rather than a reduction of the primary documentation to core metadata. The deductions are formulated to ensure the high recall of each FR by comprising the respective formulation variants of facts in the network. The FRs simulate a much simpler network of a small number of intuitive relationships, which can be combined in an "advanced search" to cover a wide range of frequent and relevant queries.

In the framework of the European Integrating Project 3D-COFORM[3], funded by the European Community's Seventh Framework Programme (FP7/2007-2013, no 231809), together with project partners we are currently implementing such a system based on the CIDOC CRM (ISO21127) [1] and extensions describing the Digital Provenance for empirical 3D modeling processes. Digital Provenance data [2] [28] form deep chains of events connected by output-input, with up to ten-thousands of intermediary products, that inherit many properties along the processing chains. We

[1] http://www.europeana.eu/portal/
[2] http://www.kulttuurisampo.fi/index.shtml
[3] http://www.3d-coform.eu/

could show that even the recall of a simple metadata element like dc:subject can much be improved by reasoning, if a richer schema is available. We could further show that many relevant queries for handling the technical data can be formulated in FRs independent from the particular domain.

2 Problem and Related Work

As described above, the only way to provide better recall than keyword search is to use suitable metadata and associative queries. We maintain that the poor performance of associative queries in Open World systems, in particular in information aggregation services, has the following reasons:

The relationship the user looked for **was not documented**, or represented **in a different way**. For instance, someone may look for things made from steel, but some objects were registered to have parts made from steel, or be of type "steel object". Our experience developing ISO21127 showed that it is impossible to normalize a global model for information integration to a unique representation for each property. Rather, in aggregation systems and the Semantic Web, one has to accept that properties are represented by reasonable alternatives that can be related by deductions. ISO21127 describes explicitly some of the most prominent alternatives as "shortcuts" (joins). The precision librarians achieve with their cataloguing rules - only after extensive training - is restricted to very small schemata (such as MARC[3]) and does not scale to the Semantic Web.

The **more analytical** and generic a global model is in the sense of formal ontologies, the **less obvious** it is for the user how a simple, intuitive question relates to the ontology. If the ontology expands very much to **application specific** and natural language properties, the user is **overwhelmed by the number of choices** and looses recall. The complexity of querying comes from properties that are transitive and cause inheritance of properties along those property paths, such as actors, place, time inherited from super- to subevents, materials from parts to wholes, subjects from a thing to its copy or derivative, narrower terms and geospatial areas inheriting broader ones, etc. Not only Digital Provenance data form such huge related graphs, but also Work-Expression-Manifestation-Item chains described by FRBR [4], see for example Rodin's 'The Gates of Hell'[4] or the movie "Bladerunner"[5] , and political history itself.

Even in our team, we were **not able** to formulate and verify such queries against ISO21127 in **SPARQL** without a more intuitive intermediate representation. Therefore, one category of systems aim to help users formulate path queries with ontology terms. This can be done with menu-guided user interfaces to specify subject-property-object triples, combined with a look-ahead enhanced search, such as realised in DBpedia[5]. It uses Virtuoso[6] for basic query-time inferences such as rdfs:subClassOf, rdfs:subPropertyOf and owl:sameAs. Other systems, like NightLite[8], employ a query formulation facility with graph representations of the ontology, but still require SPARQL knowledge. To our experience, only an expert of

[4] http://en.wikipedia.org/wiki/The_Gates_of_Hell
[5] http://en.wikipedia.org/wiki/Blade_Runner

the respective ontology can formulate queries with a reasonable recall with such systems. A layman may be impressed by reasonable answers in systems such as DPpedia, where the database is large enough, but that is not a proof of recall.

Another approach is the use of natural language queries, which are automatically mapped to associations of triples of the implemented ontology by a built-in dictionary of matching terms and synonyms and some inference mechanism, such as the Power Aqua system [7]. This approach relieves the user from learning the ontology terms, but it inherits all the well-known **polysemy of natural language**, which deteriorates precision, and provides even worse recall than the explicit use of ontology terms, because the user has no idea what can be asked or can be answered. Other natural language search systems, such as Swoogle[8] and SemSearch[9] do not interface to a triple store.

The most common approach to reduce the complexity of querying is to **reduce the complexity of the Semantic Network** itself. The Dublin Core Metadata Elements [10], VRA Core[11] and other metadata standards reduce the network to flat relationships. Another approach has been proposed by the consortium of interchange of museum information (CIMI)[6], which suggests the metadata elements: who, what, when, where ('4w') as a domain independent relations to four kinds of entities (person, thing, time, place), a kind of "faceted search". Simply, whatever relation a thing may have to a person is an answer to the question "who" etc. This works well for metadata describing only the history of objects. Otherwise, the ambiguity becomes overwhelming. Systems like Artefacts Canada[7] provide an advanced search facility based on this paradigm (plus a "how"). The method is quite intuitive and gives a good recall, but limited in application and precision.

Even if this reduction may locally improve recall, in a larger information integration environment, the lack of precision in the primary documentation, in particular the missing concept of events, results in the inability to integrate related data from different sources, as has been shown in [12], and generally in an unbearable loss of knowledge that could be rendered by the metadata. If such "simple" metadata are to be created individually for all elements of complex correlation graphs characteristic for history, interesting works of arts and e-science data, the **same facts have to be repeated** manually hundreds to ten-thousands of times, which is ineffective and error-prone, and in no ways "simple". Further, there are many relevant queries these "core fields" do not cover. All these systems **cannot be scaled up to higher precision**.

An interesting intermediate between a full-fledged semantic network and faceted search is the very successful Finnish CultureSampo[13]. It uses 9 instead of 4 facets, including material, events, object types. It uses for each facet rich term hierarchies of inclusion or subsumption, and provides multiple explicit, direct relationships between facets, such as 50 kinds of social Agent-Agent relationships. It even provides a natural language search. It comes closest to our approach. It avoids the error-prone search for suitable properties to query, provides deductions from term hierarchies, selection of valid query parameters, faceted search. It still misses other deductions than the term hierarchies, class and property subsumption.

---

[6] http://www.cni.org/projects/cimi/

[7] http://www.pro.rcip-chin.gc.ca/artefact/index-eng.jsp

As last problem, the **lack of recall in one field**, due to parameter alternatives or incompleteness of knowledge, roughly **multiplies** with the number of fields put in conjunction. A typical query in culture may ask for a particular type of things from a particular period of time, a particular geographic area, used/made by a particular group of people. If each element of the conjunction has a recall of 90%, the conjunction of four has about 66%, in the case of no correlation between the fields. As a particular problem, our query systems interpret a missing value as a negative hit, rather than as a potentially positive hit. Therefore, a successful "advanced search" facility must strive to increase recall of the individual parameters (as the 4W queries). Even better would be to indicate to the user which parameter may have been most "catastrophic" to the result.

In this paper, we propose a method that tries to combine the best from the above approaches and go beyond them.

3 Realization

The 3D-COFORM Project aims at providing integrated technologies to make the large-scale production of 3D models feasible for the systematic documentation and study of material cultural heritage. For that purpose it combines leading-edge technologies for 3D model generation from acquired data (photographic or laser), generation of synthetic models and presentation. Underlying is a scalable Repository Infrastructure to manage integrated, distributed data and metadata about cultural heritage objects, digital representations, scholarly and scientific annotations. The RI contains a metadata repository [24] we have implemented on SESAME[8] and an OWLIM[9] reasoner, that provides the platform and semantics to manage objects, 3D modeling, models and presentations likewise and supports the scholarly discourse on recent and past object features in archaeology, sites and monuments management, museum disciplines and conservation. The query system we have designed and describe in this paper is part of the "Integrated Viewer and Browser" component we are implementing together with partners. It is running on the RI.

3.1 Designing Fundamental Categories

Whereas our current implementation is based on the CIDOC CRM and extensions, our approach can be applied to other ontologies in an analogues way. However, much of its reasoning capability depends on explicit event representation, which is also present in the ABC Harmony model [17], DOLCE [18], BFO[15], Europeana EDM[23] and other ontologies. Our target domain is the generic search for things, ideas, people, and facts from the past - characteristic for Digital Libraries, cultural-historical research, science, business intelligence and political inquiries. We draw on rich previous experience in the cultural domain (such as Polemon Project[26]) and explicit queries collected from archaeologists and museum curators in 3D-COFORM.

[8] http://www.openrdf.org/
[9] http://www.ontotext.com/owlim/

In a typical Web search engine, searches would homogeneously return just Webpages, or, in a Digital Library, only documents. In a Semantic Network however, users can retrieve any instance of any class known to the system. Therefore, we firstly divide the entities of our universe of discourse into a set of relevant "Fundamental Categories" that appear to be founded deeply in our intuitive understanding of the world in this or a similar form. These FCs serve as domains and ranges of Fundamental Relationships described below. As in "core metadata", we try to cover the domain with as few FRs as possible that a user can easily learn, but still to be able to make some powerful distinctions keyword search cannot do, such as discerning places from people with the same name. In case of ambiguities, we prefer recall over precision. In the selection of the FCs, we follow the tradition of Ranganathan [14], CIMI's 4Ws and others. In our implementation, we have selected:

1. *Thing* = crm:E70.Thing[10], comprising material and immaterial things, a special case of "What" and Ranganathan's "Matter".
2. *Actor* = crm:E39.Actor, comprising persons, organisation, offices and informal groups, equal to "Who" and Ranganathan's "Personality".
3. *Event* = crm:E2.Temporal_Entity, comprising states, historical and other periods in the sense of the CRM (crm:E4.Period), and events (crm:E5.Event) and activities (crm:E7.Activity) in the narrower sense. It is equal to Ranganathan's "Energy". In some cases, periods can be regarded as a "When".
4. *Place* = crm:E53.Place, geometric extents in space, on earth and on objects, often related to or even identified by some stable and prominent configuration of matter, such as a settlement. It is equal to "Where" and Ranganathan's "Space".
5. *Time* = crm:E52.Time-Span, a date-time interval, a special case of "When" and equal to Ranganathan's "Time".
6. *Concept* = crm:E55.Type, comprising all kinds of universals, such as types of things, people, events, places, species etc. This is a special case of "What". Ranganathan and many library subject catalogues do not distinguish between particular things and types of things, however FRBR introduces the notion of "Concept".

These categories should cover the domain of interest as a "base level" distinction [25], but are neither completely disjoint nor absolute. Disjointness is actually not helpful for recall. For instance, a settlement can be at least a "Thing" and a "Place". A person (Actor) undergoing surgery, or in an excavated tomb, may be described, besides others, in terms of properties of a "Thing". This may be appear odd in other contexts. A modern biologist would regard species as "Things", i.e., human inventions with creators and other historical attributes, whereas other domains may see species only as Concepts. Therefore, the FCs should be adjustable/adjusted to the audience by adding or subtracting "less prototypical" subclasses (see [25], "prototype

---

[10] We refer to CIDOC CRM concepts here in an RDFS manner [19], denoting the CRM namespace http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.2_english_label.rdfs by "crm:"

effects), or even by extending. In the cultural-historical context, which we initially anticipated, queries with numerical values as parameters rather rare (except for dates and geo-cordinates). However, in the 3D processing domain, such queries do occur. Therefore we will add in the future "crm:E54_Dimension" to the FCs, but the generic treatment of different metrics we have not (yet) explored.

3.2 Designing Relationships

In addition to the URIs, we assign to all RDF nodes textual (non-unique) labels with names or titles. Some also have descriptions in rdf:literal form. A user formulating a query in our system may first type in a keyword. A full text search into all literals returns the associated nodes in the browser, together with minimal metadata and icons. Each node is marked by the FC it is an instance of.

For a more precise query, a user must first "select" (in the sense of the the SQL Select statement) a FC his question is about (In a normal Digital Library, this may be fixed to "document"). Then the user must compose a sort of Where Clause. The most simple one consists of a flat list of properties with range values, combined by AND or OR. The design challenge is to find a minimal set of relationships, "FRs", intuitive to the user and easy to learn, that widely cover the respective discourse with high recall and a precision enough not to be "flooded" by unrelated answers.

Fauconnier and Turner [21] observed that our subconscious maintains are much more elaborate semantic network than we are aware of, from which our conscious produces seemingly simple relations by "compression" along different dimensions, which then appear in our language. "Frames", as he calls them, of categories of constituents of respective situations allow for subconscious expansion of the meaning of attributes such as "the baby is safe", "the beach is safe", "the vacuum cleaner is safe". Following Fauconnier's research it becomes obvious that there are intuitive conscious concepts that, if turned directly into an ontology or schema - as many metadata specialists suggest - will not be suitable to support the actual reasoning humans do with these concepts. Consequently we look for selected natural expressions that can be expanded in terms of our semantic network.

Further, Pustejovsky [20] observed how language disambiguates words by the relations to other words in a phrase. For instance, "he spoke to the museum" versus "he walked around in the museum", or "he went through the door" versus "he painted the door" (from) seems contradictory in an ontology, but do not surprise people in whatever language we translate it to. This "complementary polysemy ", as he calls it, can be explained by classifying contextual expressions into relatively few, language-neutral categories ("quales"). When a user selects a relationship term and a value, we use a similar mechanism to disambiguate the term as a further help to the user: The term is interpreted according to the selected FC and the FC the value is instance of, rather than forbidding "illegal values". Of course the user may also filter values by the FC.

A good example is the term *from*, a very natural relationship term describing a sort of origin or provenance. For instance, in good museum practice and intuition "Things *from* New Guinea" may mean things found, produced, or used in New Guinea or things with parts from there. It may also mean things produced by people coming from New Guinea. This interpretation is common for all Place values. Museum metadata frequently contain the term “provenance” in this sense. However, “Things *from* J.W. Goethe” (an Actor) has a different interpretation: It could mean things created, produced, modified, said, acquired, owned, kept or used by him or his household, gifts he gave or received, or awards he received. “Things *from* the Parthenon” (a Thing) may mean parts or pieces of the Parthenon, but it may also comprise inscriptions found on it. Quite differently, we would interpret “Actors (people) *from* New Guinea”, a sort of nationality concept, whereas “Actors (people) *from* Siemens Company” (Actor) would pertain to membership. “Places *from* Time” make no sense. All interpretations correspond to composite path expressions in the CIDOC CRM. Constrained to a particular combination of FCs, it is feasible to find all relevant expressions in the ontology for this interpretation.

Our empirical sources for the FR are “simple” metadata schemata, such as Dublin Core and VRA, but also the Europeana EDM model, experiences from structuring museum information [22], generalizations of the CRM itself and intuition. We divide the relationships into those describing (1) how and what something is (classification, part-whole structure), (2) what an item has undergone gone in its history, and (3) what it may “show”, say or refer to. We have not looked at relationships of intention, motivation or cause, because they are rarely documented. In our current implementation, we have selected:

1. *has type:* denotes relations of an item[11] to a classification, category, type, essential role or other unary property, such as a format, material, color. It generalizes over dc:type, dc:classification, dc:format, dc:language. The relationship is applicable to all FCs and has always range Concept.
2. *is part of*: denotes structural relations of an item to a wider unit it is contained in. The relationship is applicable to all FCs, except for Concept. In case of Actors, one would rather speak of “*is member of*”, and persons are the minimal elements. Domain and range must be identical.
3. *is similar or the same with*: denotes the symmetric relation between items that share features or are possibly identical. It is only usual for Things to document similarity manually. There exist enough comparison algorithms that deduce degrees of similarity automatically. We do not deal with these in this work.
4. *has met:* denotes the symmetric relation between items that were present in the same event, including time intervals and places. Applicable to any combination of FCs, except for Concepts.
5. *from, has founder* or *has parent:* denotes the relations of an item to constituants of a context in its history which is either significant for the item, or the item is significant for the context, “provenance” in the widest sense, including time intervals and places. In case of genealogy or group formation, natural language prefers the terms parent and founder respectively in order to refer to Actors. The relationship is a special case of *has met*. The applicability is analyzed in Table 1.

[11] By “item” we mean any instance of a Fundamental Category, similar to “resource” in RDF terms.

6. *is origin of, founder of, parent of*: the inverse of *from, has founder* or *parent*. In case of Actor as domain, one would rather speak of ***"is provider or creator of "***
7. *at:* denotes the relation of an Event to the Time and Place within which the event happened.
8. *refers to:* denotes the relation of an item that is information, contains information or has produced information to the item this information refers to or is about. The relation can even be extended to a Place from where such information originated.
9. *is referred by/ is referred to at:* the inverse of *refers to.*

Table 1 describes which of the above relationships are applicable to respective combinations of FCs as domain and range. Each relationship has a different interpretation for each applicable combination of domain and range, which adapts the general meaning described above to the concrete case, as explained above.

Table 1. Fundamental Categories and Fundamental Relationships.

Domain (select)	Range(query parameter)				
	Thing	Actor	Place	Event	Time
Thing	2.is part of 3.is similar or the same with 4. has met 5. from 6. is origin of 8. refers to 9.is referred by	4.has met 5.from 8.refers to 9.is referred by	4.from 8.refers to 9.is referred to at	4.from 8.refers to	4.from
Actor	4.has met 6.is creator or provider of 8. refers to 9.is referred by	2.is member of 4. has met 5.has parent or founder 6.is parent or founder of 8.refers to 9.is referred by	4.has met 5.from 8.refers to 9.is referred to at	4.has met 8.refers to 6.has met	8.refers to 6.has met 4.from
Place	5.is origin of 8.refers to or is about 9.isreferred by	5.is origin of 8.refers to or is about 9.is referred by	2.is part of 5.is origin of	9.is referred by 5.is origin of	7.at
Event	5.is origin of 9.is referred by 8.refers to or is about	4.from 9.is referred by 8.refers to or is about 6.has met	8.refers to or is about 7.at	8.refers to or is about 2.is part of	8.refers to or is about 7.at
Time	5.is origin of	5.is origin of	5.is origin of	5.is origin of	2.is part of

The category Concept plays a special role. Concepts can be subdivided into subtypes of the FCs themselves, such as "Thing Concepts", "Place Concepts", etc.

Relation 1. *has type* has domain all FCs, but the range is restricted to subtypes of the domain: Thing. *has type:* Thing-Concepts, etc. Further, all relationships in table 1 can be extended into categorical questions: For instance the relation “Things *from* Place” expands into Things *from type of* Place” via a join with “Place *has type* Place Concept”. This is implemented as generic mechanism. The relationships in table 1 are not all disjoint. There are some subsumption relations between them, for instance *has met* is more general than *from.*

Our framework foresees open-ended specialization of the FRs to dynamically meet increased precision demands. For instance, Thing *was created at* Place is an obvious specialization of Thing *from* Place. The user will be able to browse from the FRs to their specializations. The user interface will further allow the user to combine the FRs even to simple path expressions, as if they were properties of the network, for instance “all things *from* events *of type* X *at* a time Y and *at* a place Z”. Such a customizable system of predefined where clauses that appeared as properties that can be combined to other queries we had already implemented in [26].

3.1 Experimentation

So far we have expanded most of the relationships in table 1 into respective queries to the CIDOC CRM and CRM Digital, and now are in the process of verifying them one by one against real data and with users in 3D-COFORM. The process is quite time-consuming and takes months. Systematically, all “simple” metadata relationships give rise deduction for better recall. Properties are inherited or transferred along (1) derivation chains, (2) from parts to wholes and wholes to parts and (3) induced by alternative processes or descriptions of varying detail. See for instance “Thing *from* Place” in the Appendix. Even “dc:creator” is frequently an indirect property: for instance, Rodin has never seen Rodin’s bronze statue “Monument to Balzac”. In general, it is impossible to expand all those deductions in the primary documentation because of the scale and limitations of local knowledge. It is equally impractical to expect the end-user to redetect all those relevant paths expressions at query time.

Due to the rigorous generalization of classes and properties in the CRM, we found it feasible to enumerate all alternative deduction paths. For instance, creation, production, and other processes are subsumed under “crm:E63. Beginning_of_Existence”. Many of the deductions actually describe the **probability or fuzzy condition** that the respective property is inherited or transferred. For instance, if a thing which has parts from some country is regarded to be from this country may vary from case to case or user to user. So, even though we query a network of logical associations, some deductions acquire a fuzzy or probabilistic nature, which brings us, luckily or sadly, back to the old recall-precision optimization problem known well enough from information retrieval. The most complex case we could demonstrate so far was the query for “all things that *refer to* Ivory Panel A.15-1955”, an object from the Victoria & Albert Museum digitized twice, which comprised dozens of processing steps and intermediate files. The query collected all data files showing the object, even though the fact was only deduced from the existence of the “Digitization Processes” in the Provenance chain. The full deduction

"Thing refers to Thing" comprised 6 alternative paths including 8 transitive closures. The object itself has subject the "Ascension, Christ", so we can even retrieve those files under "all things that refer to Christ".

5 Conclusions and Further Work

We prose a new framework for querying semantic networks: For formulating queries, the user is presented a small list of configurable "Fundamental Relationships" and relevant specializations, easy to comprehend, that abstract by rich deductions from an underlying semantic network of much more specialized metadata comprising explicit event descriptions. These FRs simulate to the user a much simpler semantic network, which covers as many generic questions as possible with a high recall. The specializations of the FRs allow for systematically increasing the precision of queries on demand, down to the level of detail of the underlying network.

With this method, we believe we can overcome the recall-precision gap between keyword and semantic search, the problems of formulating powerful queries in complex semantic networks and the problems of simplifying the metadata themselves, but, of course, rely on an efficient database technology. Future work will consist of further testing, consolidating and refining the FRs with respect to real user questions, including practical 3D data management and scholarly queries. It is planned to upload complete museum collection data to the RI and to deploy it for massive 3D model production, but other large-scale information integrators may take up the method as well. The complete analysis of the FRs will be published in a technical paper after sufficient testing.

References

1. Doerr M., The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata, AI Magazine, Volume.24, Number 3, pp.75--92 (2003) , http://www.cidoc-crm.org/index.html
2. Theodoridou M., Tzitzikas Y., Doerr M., Marketakis Y., Melessanakis V.: Modeling and Querying Provenance by Extending CIDOC CRM. Distributed and Parallel Databases (2010).
3. MARCXML Schema, http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd
4. O'Neill, E., T., FRBR: Functional Requirements for Bibliographic Records, Application of the Entity-Relationship Model to Humphry,(2002)
5. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.. DBpedia: a nucleus for a web of open data. In: 6th international The semantic web and 2nd Asian conference on Asian semantic web conference, pp.722--735. Springer, Heidelberg (2007)
6. Virtuoso: Inference Rules & Reasoning, http://docs.openlinksw.com/virtuoso/rdfsparqlrule.html

7. Lopez, V., Motta, E., Uren, V.. PowerAqua: Fishing the Semantic Web. In: Sure, Y., Domingue, J. (eds.) The Semantic Web: research and Applications. LNCS, vol. 4011, pp. 393--410. Springer, Heidelberg (2006), http://technologies.kmi.open.ac.uk/poweraqua/
8. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V., Sachs, J.. Swoogle: a search and metadata engine for the semantic web. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management, pp. 652--659.ACM New York, NY 2004
9. Lei Y., Uren V., Motta E. SemSearch: A Search Engine for the Semantic Web. In: Staab, S., Svátek, V. (eds.) Managing Knowledge in a World of Networks. LNCS, vol. 4248, pp. 238--245. Springer, Heidelberg (2006)
10. Dublin Core Metadata Initiative, http://dublincore.org/documents/usageguide/elements.shtml
11. VRA Core 4.0 Element Description, http://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf
12. Doerr, M., Iorizzo, D. The dream of a global knowledge network-A new approach. J.Comput.Cult.Herit.1 (1), 1-23 (2008)
13. Hyvoenen, E., Ruotsalo, T., Haeggstroem, T., Salminen, M., Junnila, M., Virkkilae, M., Haaramo, M., Maekelae, E., Kauppinen, T., Viljanen, K.: CultureSampo—Finnish Culture on the SemanticWeb: The Vision and First Results. In:12th Finnish Artificial Intelligence Conference STeP (2006). http://www.kulttuurisampo.fi/index.shtml
14. Ranganathan, S.R.. A descriptive Account of Colon Classification", Bangalore: Sarada Ranganathan Endowment for Library Science (1965)
15. Basic Formal Ontology, http://www.ifomis.org/bfo
16. Day, M., Duval, E., Jong, A., Sfeir, E., Causton, L., Koch, W., Baker, T., Heery, R., Patel, M.. Metadata Watch Report #3, PricewaterhouseCoopers (2000)
17. Lagoze, C., Hunter, J. The ABC Ontology and Model, Journal of Digital Information (2001)
18. Gangemi A., Guarino N., Masolo C., Oltramari, A.,Schneider L. 2002. Sweetening Ontologies with DOLCE. In Proceedings of EKAW 2002, Siguenza, Spain, pp 166-178.
19. McBride, B., The resource description framework (RDF) and its vocabulary description language RDFS, Handbook on Ontologies, pp.51-66, (2004)
20. Pustejovsky, J.: The generative Lexicon. Comp. Ling. 17 (1991)
21. Fauconnier, G. and Turner, M., The way we think: Conceptual blending and the mind's hidden complexities, Basic Books, New York (2002)
22. Bekiari, C. and Charami, L. and Doerr, M. and Georgis, C and Kritsotaki, A.. Documenting Cultural Heritage in Small Muesums. In: 2008 Annual Conference of CIDOC (2008)
23. Doerr, M. and Gradmann, S. and Hennicke, S. and Isaac, A. and Meghini, C. and van de Sompel, H.: The Europeana Data Model (EDM). In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY (2010)
24. Doerr, M., Tzompanaki, K., Theodoridou, M., Georgis, Ch., Axaridou, A., & Havemann, S.: A Repository for 3D Model Production and Interpretation in Culture and Beyond. In: VAST 2010: 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage. Palais du Louvre, Paris, France. pp. 97-104. (2010)
25. Lakoff, G.: Women, fire, and dangerous things: What categories reveal about the mind.University of Chicago Press (1987)
26. Bekiari, Ch., Gritzapi, Ch.D., & Kalomoirakis.: POLEMON: A Federated Database Management System for the Documentation, Management and Promotion of Cultural

Heritage. In: 26th Conference on Computer Applications in Archaeology. Barchelona, Spain (1998)
27. Fernandez, M., Lopez, V., Sabou, M., Uren, V., Vallet, D., Motta, E., Castells, P.: Semantic Search meets the Web. In: The IEEE International Conference on Semantic Computing. pp 253-260 (2008)
28. L. Moreau, J. Freire, J. Myers, J. Futrelle, and P. Paulson. The Open Provenance Model. University of Southampton, 2007.

6 Appendix

As an example we describe the deduction **Thing *from* Place.** The domain, instance of Thing, is denoted by "X", and the range value, instance of Place, by "Y". By superscripts (0,n) we denote a transitive closure of a property, including not applying it at all. Between domain and property we put a dot, and a colon between property and range. In terms of the CIDOC CRM namespace http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.2_english_label.rdfs, the deduced relationship is:

X.((P46F.is_composed_of$^{(0,n)}$) OR (P106F.is_composed_of$^{(0,n)}$) OR (P148F.has_component$^{(0,n)}$)):((E70.Thing.P92B.was_brought_into_existence_by:E63. Beginning_of_Existence.P9B.forms_part_of$^{(0,n)}$:E63.Beginning_of_Existence.P7F.took_place_at:**Y**) OR (E70.Thing.P16B.was_used_for:E7.Activity.P9B.forms_part_of$^{(0,n)}$:E7.Activity.P7F.took_place_at:**Y**) OR (E24.Physical_Man-Made_Thing.P128F.carries:E73.Information_Object.P94B.was_created_by:E65.Creation. P7F.took_place_at:**Y**) OR (E70.Thing.P92B.was_brought_into_existence_by**:** E7.Activity.P14F.carried_out_by:((E39.Actor.P107B.is_current_or_former_member_of$^{(0,n)}$:E74.Group.P74F.has_current_or_former_residence:**Y**) OR (E21.Person.P98B.was_born:E67.Birth.P7F.took_place_at:**Y**)))