D4.2 DATA MODELS AND ALIGNMENT

HTTP://WWW.MOLTO-PROJECT.EU

Contract No.:	FP7-ICT-247914
Project full title:	MOLTO - Multilingual Online Translation
Deliverable:	D4.2 Data Models and Alignment
Security (distribution level):	Public
Contractual date of delivery:	30.05.2010
Actual date of delivery:	16.05.2010
Type:	Report
Status & version:	Final
Author(s):	Mariana Damova
Task responsible:	ONTO (WP4)
Other contributors:	Dana Dannélls, Inari Listenmaa

ABSTRACT

THIS DOCUMENT DESCRIBES THE DATA AND THE DATA MODELS TO BE INCLUDED IN THE KNOWLEDGE REPRESENTATION INFRASTRUCTURE (KRI) DESCRIBED IN DELIVERABLE 4.1. THE ADOPTED APPROACH OF LINKED DATA IS EXTENDED TO MULTILINGUAL COVERAGE, AND APPLIED TO DATA FROM THE CULTURAL HERITAGE SUBJECT DOMAIN. THIS DOCUMENT PRESENTS THE METHODS OF INTERLINKING THE DIFFERENT RESOURCES AT SCHEMA AND AT INSTANCE LEVEL, AND PROVIDES INSIGHT ABOUT THE INTEGRATION OF NEW DATA INTO THE KRI.

CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

The knowledge representation infrastructure (KRI) of MOLTO is a collection of technologies enabling interoperability between natural language representations and structured knowledge modeled in RDF - conceptual models (ontologies) and knowledge bases (instances). The main purpose of this is to ensure natural and easier access to knowledge through semantic search, exploration and knowledge discovery. To make this possible, the KRI has to be filled in with actual content, data. Having the knowledge engineering infrastructure in place, we focus on developing the conceptual models and knowledge bases needed to realize the grammar development (WP2) and the use cases of MOLTO (WP6-8). This document is dedicated to describing the development of the data part of the KRI. This includes the data models, the instance data integration and the setting up of the KRI data environment.

In the last years, a rapidly increasing amount of various datasets has been made available in a machine readable form, through W3C standards like the Resource Description Framework (RDF) [34], the Web Ontology Language (OWL) [32] and initiatives like Linked Open Data (LOD) [25]. We adopt the approach of Linked Open Data to construct the knowledge base of KRI. RDF(S) and OWL have been developed to formally express the semantics of data and content, serving as the fundaments of the emerging Semantic Web. Both RDF(S) and OWL have been initially designed to express formal meaning representations of data and content in machine readable form.

In this deliverable we outline the criteria for selection of a representative collection of datasets and describe the methods of interconnecting them which allow to obtain a very rich pool of factual knowledge, based on explicit and implicit facts, produced from materialization in the time of data loading. LOD alone points to close to four hundred datasets, aligned between themselves, capturing various subject areas, from DBpedia, the Wikipedia structured exports, through FOAF profiles, thesauri like WordNet, movie and music databases, and all the major scientific bio-medical datasets. The approach to develop KRI data is based on LOD approach, and extended by a method of using intermediary layers of upper-level ontologies which provide a unified way to access the pool of heterogeneous data. The base of this is PROTON ontology, linked to the DBpedia ontology which together with its instances forms a large coverage knowledge base of common sense knowledge. The applicability of the adopted approach of building the KRI knowledge base to the use cases of MOLTO is demonstrated by showing how to integrate the cultural heritage ontology - CRM (Conceptual Reference Model) into the KRI data, and how to populate the KRI knowledge base with real data from the Gothenburg museum.

Additionally, it is shown how the KRI knowledge base provides access to lexical knowledge to be used in the process of treating the natural language input by the GF grammar rules, cf. WP2, and provide the basis for multilingual support. This is done in this deliverable by showing how a bilingual English–Finnish variant of the world's largest lexical knowledge base WordNet is integrated into the KRI knowledge base.

This deliverable outlines the different knowledge sources and data models necessary for making the KRI operational in the context of MOLTO project. It is organized in the following way: Sections 2 and 3 describe the principles of Linked Open Data, and a method to successfully manage them. Section 4 gives the methods of selecting the basic KRI knowledge base, and section 5 outlines the data integration approaches for KRI. Sections 6 and 7 introduce the data necessary to handle the museum use case, e.g. the CRM ontology and the data from the Gothenburg museum together with the methods of extending the KRI knowledge base with them. Section 8 turns to WordNet and the integration of its bilingual (Finnish-English) version into KRI knowledge base. Section 9 and 10 discuss the relationship between natural language grammars and ontologies. Section 11 is about the experimental environment and the experiments to be performed with KRI and with the constructed KRI knowledge base. Section 12 concludes the deliverable.

The intended readership of this work includes scholars, knowledge engineers, developers and CIOs interested in linked data management, ontology engineering, information infrastructures, data integration, semantic technologies, interoperability between natural language and general knowledge bases.

2 LINKED OPEN DATA - THE VISION

The notion of "linked data" is defined by Tim Berners-Lee, [1], as RDF graphs, published on the WWW so that one (machine or human) can explore them across servers by following the links in the graph in a manner similar to the way the HTML web is navigated. It is viewed as a method for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF. "Linked data" are constituted by publishing and interlinking open data sources, according to the following principles:

1. Using URIs as names for things;
2. Using HTTP URIs, so that people can look up those names;
3. Providing useful information when someone looks up a URI;
4. Including links to other URI, so people can discover more things.

In fact, most of the RDF datasets fulfil principles 1, 2, and 4 by design. The piece of novelty in the design principles above concerns the requirement for enabling Semantic Web browsers to load HTTP descriptions of RDF resources based on their URIs. To this end, data publishers should make sure that:

- the "physical" addresses of the published pieces of data are the same as the "logical" addresses, used as RDF identifiers (URIs);
- upon receiving an HTTP request, the server should return a set of triples that describe the resource.

Linking Open Data (LOD) is a W3C SWEO community project aiming to extend the Web by publishing open datasets as RDF and by creating RDF links between data items from different data sources. Linked Open Data provides sets of referenceable, semantically interlinked resources with defined meaning. The central dataset of the LOD is DBPedia. Because of the many mappings between other LOD datasets and DBPedia, the latter serves as a sort of a hub in the LOD graph assuring a certain level of connectivity. LOD is rapidly growing – as of September 2011 it contains close to 400 datasets, with total volume above 30 billion statements, interlinked with 395 million statements, cf. Figure 1.

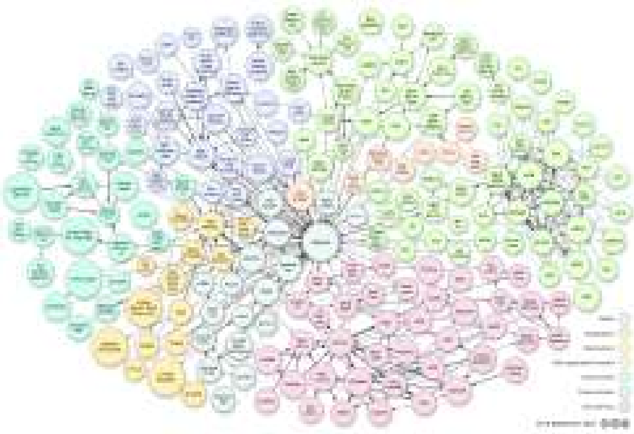

Figure 1

The use of LOD and the development of applications based on it is difficult because the different LOD datasets are rather loosely connected chunks of information, facts, instances. Although all of them have schemas describing their structure, the linkage between the datasets in LOD is predominantly made at the level of single facts, thus missing the potential of aligning them at the schema level. The schema-level alignment of LOD datasets presents serious advantages related to ensuring the consistency of the linkages. Such linkages can enable applications that can answer queries requiring multiple and disparate information sources.

3 LINKED OPEN DATA MANAGEMENT

Using linked data for data management is considered very promising and with great potential. However, there are several challenges coming from the current state of the LOD cloud that have to be handled before making this possible. These are summarized in the following points:

- LOD is hard to comprehend – the fact that multiple datasets are interlinked and accessible in the same data format does not on itself help in dealing with hundreds of data schemata, ontologies, vocabularies and data modelling patterns;
- Diversity comes at a price – often there are tens of different ways of expressing one and the same information in even in a single dataset such as DBPedia;
- LOD is unreliable – many of the servers behind LOD today are slow and have down times higher than the one acceptable for most of the data management setups;
- Dealing with data distributed on the web is slow – a federated SPARQL query that uses, say, 3 servers within several joins can be very slow;
- No kind of consistency is guaranteed – low commitment to the formal semantics and intended usage of the ontologies and schemata.

Reason-able views, [23], represent an approach for reasoning with and management of linked data defined at Ontotext. A *reason-able view* is an assembly of independent datasets, which can be used as a single body of knowledge with respect to reasoning and query evaluation. The key principles constituting the *reason-able views* are as follows:

- Group selected datasets and ontologies in a compound dataset;
- Clean up, post-process and enrich the datasets if necessary. Do this conservatively, in a clearly documented and automated manner, so that (i) the operation can easily be performed each time a new version of one of the datasets is published and (ii) the users can easily understand the intervention made to the original dataset;
- Load the compound dataset in a single semantic repository and perform inference with respect to tractable OWL dialects;
- Define a set of sample queries against the compound dataset. These determine the “level of service” or the “scope of consistency” contract offered by the reason-able view.

Each *reason-able view* is aiming at lowering the cost and the risks of using specific linked data datasets for specific purposes. Each *reason-able view* follows specific design objectives that are:

- Make reasoning and query evaluation feasible;
- Lower the cost of entry through interactive user interfaces and retrieval methods such as URI auto-completion and *RDF search*(a search modality where RDF molecules are being retrieved and ranked by relevance to a full-text style query, represented as set of keywords);
- Guarantee a basic level of consistency – the sample queries guarantee the consistency of the data in the same way in which regression tests do for the quality of software;
- Guarantee availability – in the same way in which web search engines are usually more reliable than most of the web sites; they also do caching;
- Easier exploration and querying of unseen data– sample queries provide re-usable extraction patterns, which reduce the time for acquaintance with the datasets and their interconnections.

The *reason-able views* are built according to the following design principles:

- all datasets in the view represent linked data,
- single reasonability criteria is imposed on all datasets,
- each dataset is connected to at least one of the others.

Reason-able views are implemented in two systems at Ontotext, namely, FactForge (http://factforge.net) and LinkedLifeData (http://www.linkedlifedata.com), and are integrated in the semantic knowledge base of a large European archive containing government data and documents.

The single repository in which the datasets are loaded is Owlim [2]- one of the most scalable triple stores in the world with a very fast loading and querying parameters. Full materialization is performed during loading based on Owl-Horst inference [33] and some processing optimizations, described in detail in [22]. As a result of this, the total number of available for querying RDF statements is usually about 40% higher than the number of explicitly introduced ones.

The *reason-able views* are accessible via a SPARQL end point and keywords. Being compound datasets they give the opportunity to formulate queries combining predicates from different datasets and ontologies in a single SPARQL query. The results from such queries return instances which also come from different datasets in the compound dataset. For instance the query about the entertainers born in Germany, given below, uses information from 4 datasets of FactForge, including two ontologies and returns information from DBPedia, cf. Figure 2.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbp-ont: <http://dbpedia.org/ontology/>
PREFIX dbpedia: <http://dbpedia.org/resource/>
PREFIX opencyc: <http://sw.opencyc.org/2008/06/10/concept/en/>
PREFIX geo-ont: <http://www.geonames.org/ontology#>
PREFIX om: <http://www.ontotext.com/owlim/>

SELECT *
WHERE {
     ?Person dbp-ont:birthPlace ?BirthPlace ;
             rdf:type opencyc:Entertainer ;
             om:hasRDFRank ?RR .
     ?BirthPlace geo-ont:parentFeature dbpedia:Bavaria .
} ORDER BY DESC(?RR) LIMIT 100
```

Figure 2

MOLTO KRI knowledge base is a reason-able view of the web of data. Its design principles are described in the following sections.

4 CRITERIA FOR SELECTION OF DATASETS

The datasets in each reason-able view depend on the underlying purpose of use of the compound dataset. For instance, LinkedLifeData aims at providing access to the Life Science part of the LOD cloud. It gathers more than 20 LOD cloud datasets and handles more than 5 billion RDF statements. FactForge on the other hand aims at providing access to heterogeneous knowledge, including general purpose knowledge, linguistic knowledge, geographic knowledge. It is the biggest body of heterogeneous knowledge on which inference has been performed. It handles about 2 billion RDF statements gathering 8 of the most central LOD datasets.

The KRI reason-able view has to be constructed to provide adequate content for the requirements for:

- ability to recognize named entities, such as people, institutions, locations
- ability to handle specific subject domains, such as the cultural heritage and museums
- ability to ensure linguistic support, such as lexical knowledge bases, and multilingualism

Thus, the KRI reason-able view comprises a heterogeneous dataset reflecting a combination of generic knowledge, domain specific knowledge, and linguistic knowledge. It includes the following datasets from the LOD cloud:

- DBPedia [11] - the RDF-ized version of Wikipedia, comprising the information from Wikipedia infoboxes, designed and developed to provide as full as possible coverage of the factual knowledge that can be extracted from Wikipedia with a high level of precision. DBPedia describes more than 3.5 million things and covers 97 languages. 1.67 million of DBPedia things are classified in a consistent Ontology, including 364,000 persons, 462,000 places, 99,000 music albums.

The DBpedia knowledge base has over 672 million RDF triples out of which 286 million extracted from the English edition of Wikipedia and 386 million extracted from other language editions.

- **Geonames** [18] - a geographic database that covers 6 million of the most significant geographical features on Earth organized in a concise ontology of about 30 concepts and properties, and contains over 8 million geographical names and consists of 7 million unique features whereof 2.6 million populated places and 2.8 million alternate names, integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat/long coordinates are in WGS84 (World Geodetic System 1984).
- **WordNet** [28] - the biggest lexical knowledge base, to be described in detail in section 8 below.

These datasets cover the generic knowledge and the linguistic part of the requirements, and present an acceptable volume of data in the compound dataset amounting to several millions of RDF triples.

DBpedia and Geonames are interlinked together at the level of facts using the predicate `owl:sameAs`. The way DBpedia and WordNet are interlinked is described in section 8 below.

It has been pointed out in [21] that the reason-able views provide a practical compromise between the open-world assumption of the web and the reasonability of LOD providing a viable approach for the management of these data. However, using the reasonable views in their “raw form” causes difficulties. One, both human or computer agent, has to be aware of all predicates and all distinct schemata of the datasets which make part of the reason-able view. In many cases a number of predicates practically denote the same object, or the same concept. The schemata are often designed bottom up, and cover the data that are introduced in the datasets without respecting principles of balanced ontology design, depth, and in some cases consistency of the data. In the case of DBPedia, which is the hub of LOD, its ontology covers only a part of the data available in DBPedia, the rest of them is covered by properties, which are not organized in a structured way, but often need to be used when trying to access particular type of information.

This leads to the idea of introducing intermediary layers of schemata – upper-level ontologies –, which will ensure consistency and ease of access of the datasets of the reason-able view. The following describes the implementation of this idea within FactForge, by adopting two upper-level ontologies with different degrees of generality – PROTON [39] of about 550 concepts and UMBEL [40] of about 28000 concepts. They are aligned uni-directionally, using subsumption relations from the more specific to the more generic concept.

5 DATA INTEGRATION APPROACHES

Ontology matching is a key interoperability enabler for the Semantic Web, as well as a useful tactic in some classical data integration tasks. It refers to the activity of finding or discovering relationships or correspondences between entities of different ontologies or ontology modules. Matching ontologies enables the knowledge and data expressed in the matched ontologies to interoperate. Different methods of ontology matching have been developed in the last couple of years. There are syntactic and semantic matching systems [14]. In the syntactic matching the relations are computed between labels at nodes, and they are evaluated as [0, 1]. In the semantic matching the relations are computed between concepts at nodes, and they are evaluated as set theoretic relations. The semantic matching discovers semantic relationships across distinct and autonomous generic structures and recognizes relationships between matched entities, such as equivalence, subsumption, disjointness and intersection. When integrating two models, substantial difficulties may arise in transforming information from one model to the other in a heterogeneous context. Harmonising semantics is one approach for model integration by formal mapping between two domains. In this approach reference ontology is built to provide the link between the two models [14]. Except for the types of relationships that are matched between the ontologies, distinctions are made in the way the two initial ontologies can be accessed. Thus, there are bidirectional and unidirectional matching methods. The bidirectional method ensures access to the two ontologies from the two ontologies, whereas the unidirectional method ensures access from one to the other ontology only [14]. Another difference in the matching methods is in the way the matching is done. There is manual and automated matching. Automated matching is suitable for simple ontologies and simple matching tasks, where the exact accuracy of the matching is not of highest importance. In automatic matching structures that are being matched are labeled with natural language typically using WordNet – vocabulary mapping. It consists in comparing Classes, Properties and Instances of two ontologies in a relation one to one. The main drawback of automated ontology matching systems is that they cannot cope with ontological heterogeneity. The fact is ignored that the classes and the properties may be described in different unrelated ontologies, thus the algorithms cannot discover hidden relationships that hold between unrelated entities. Mapping by hand is considered difficult, time consuming and too long, but it derives the most accurate results. Manual mapping is suitable when maximum quality of mapping is sought for a small quantity of concepts. This is the case with the mapping of PROTON to DBpedia and Geonames schema. UMBEL is also mapped to DBpedia schema and DBpedia instances are assigned UMBEL classes. PROTON is mapped to UMBEL schema, thus ensuring two leveraged intermediary layers to the datasets of LOD.

These intermediary layers, reference ontologies, provide unified access points to the LOD datasets that are part of the reason-able view, and ensure better interoperability between them. The access to the data takes place initially via the predicates of the most generic ontology, in our concrete case PROTON, and when its predicates are not sufficient, the more specific layer UMBEL is used to retrieve data based on more specific criteria.

The described in the following mapping methods present the methodology that is to be followed when adding new datasets and new ontologies into the reason-able view. There are manual mapping techniques, which cover structural differences, that require definition analysis and also semi-automatic mapping techniques that rely on linguistic analysis of the textual parts of the definitions of the ontological entities.

Mapping PROTON to DBPedia consists in matching of ontologies and data structures built according to different methods and design principles, e.g. data-driven ontologies and an upper-level ontology. DBPedia ontology includes many ad hoc predicates, which appear in only one or several statements reflecting the variety of knowledge included in it. It is shallow and counts 24 first level concepts of very different degree of generality ranging from the philosophical concept of “event” through “person” and “place” to very specific concepts like “beverage”, “drug”, “protein”. Except for the ontology, DBPedia has a compendium of concepts and properties gathered in a separate namespace, which pertain to ontological dimensions, but are not modelled in the ontology. These concepts are also considered in the process of mapping as they give access to more data in DBPedia. The differences in the design methods of the ontologies to be mapped show cases of structural differences and hidden relations. For example, Figure 3 shows the difference in the conceptualization of the professions in DBPedia and in PROTON.

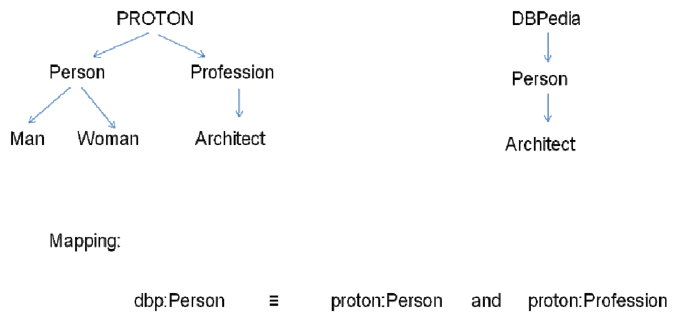

Figure 3

DBPedia describes the professions as a subclass of the class Person, e.g. refers to the profession with the professional person, whereas PROTON separates the Person from the profession he has, and links the two entities with the property `hasProfession` with `domain:Person` and range:Profession. This requires the mapping rule to be between a concept and an expression, e.g.

```
dbpedia:Professional rdfs:subClassOf [ rdf:type owl:Restriction ;
                                       owl:onProperty proton:hasProfession ;
                                       owl:hasValue proton:Professional ] .
```

The next two examples point to matching methods employed to match predicates from other LOD datasets, e.g. Freebase and Geonames, and are given here to complete the spectrum of techniques necessary to obtain matching with 100% precision between ontologies built according to different design principles.

Except for using of expressions in the mapping rules, the structural mismatches between the data-driven and the upper-level ontology trigger the need of generating new instances in the knowledge base to fill in the gaps in the diverging representations. For instance, Freebase conceptualizes the ownership relation in two steps, as a relation between (a) the object and an act of owning, and (b) the act of owning and a given object, whereas PROTON has a direct predicate linking an owner to the owned object (cf. Figure 4).

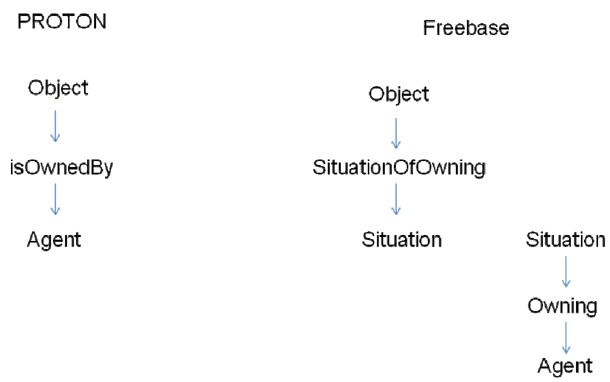

Figure 4

This enrichment is specific to the OWLIM repository and takes place at the time of loading of the knowledge base into the repository. Adding inference rules to its standard owl-max ruleset enables it. An example of such a rule, handling the discussed mismatch in the representation of the ownership relation, is shown below:

```
Id:owner
  a <fb:visual_art.artwork.owners> b
  b <fb:visual_art.artwork_owner_relationship.owner> c
  ------------------------------------------------------
  a <ptop:isOwnedBy> c
```

It generates the additional fact a <ptop:isOwnedBy> c based on the availability of the combination of the facts of the premiss.

The need of optimal coverage of the data described by the data-driven ontology drives the extension of the upper-level ontology. Thus, before drafting the mapping rules between single concepts, new concepts are to be introduced into PROTON. For example, Geonames has a concept Map, which refers to a webpage of a map. Such a concept does not exist in PROTON, whereas other types of documents like Report, Contract, etc. do exist. So, abiding to the design principles of completeness and balancing of the upper-ontology, the concept Map has been introduced to PROTON.

To ensure the unidirectionality of the mappings, which guarantee the accessibility of the data through the more general ontology first, all mapping rules between FactForge and PROTON embody subsumption relations where the FactForge concepts are considered more specific than the PROTON ones, e.g.

```
FactForge:Concept rdfs:subClassOf PROTON:Concept .
FactForge:Property rdfs:subPropertyOf PROTON:Property .
```

The alignment of PROTON with DBPedia ontology, Geonames ontology and Freebase has been hand crafted as this has been estimated as the most suitable approach to find the exact correspondences of the small amount of upper-level concepts with 100% precision and capturing of hidden relationships. The matching of the concepts and properties between DBPedia and PROTON, between Geonames and PROTON and between Freebase and PROTON has been based on comparing the definitions of the concepts, their use and instances.

As a result of the mapping processes a new version of PROTON (PROTON 3.0) has been created which has 542 classes, an increase of about 40% with respect to its original version, 128 Object properties and 55 Datatype properties. The mappings cover 203 mapped classes to DBPedia; 23 mapped properties to DBPedia; 6 mapped classes to Geonames; 14 mapped properties to Geonames; 382 mapped Geonames Codes; 9 mapped classes to Freebase; and 60 mapped properties to Freebase.

The mappings between UMBEL and DBPedia are geared to allow any of the constituent ontologies or their predicates to be used in conjunction with the other ontologies with the purpose UMBEL to assume the role of a central reference ontology for the LOD cloud. UMBEL at the level of reference concepts has been mapped to DBPedia vocabulary by hand. 272 DBpedia ontology classes are directly mapped to corresponding UMBEL reference concepts.

These class mappings then become the basis for manual or semi-automatic mappings to Wikipedia instances (pages) using either the:

- DBpedia Ontology;
- Semantic Vectors 0 [35] correspondences; or
- Analysis of DBpedia category structure.

As a result, the number of UMBEL reference concepts expanded from 20,512 to 27,917. These are all fully integrated into the UMBEL ontology with one of the proper to UMBEL 33 SuperTypes (ST) assigned.

Across all mappings, 3,527 UMBEL reference concepts are linked directly to Wikipedia (DBpedia). The result is that 2,130,021 unique DBpedia pages are in total linked to this structure via nearly 4 million predicate relations (3 935 148). All of these pages are also characterized by one or more STs.

Of these 2 million pages, 876,125 are assigned a specific SuperType via `rdf:type`; the remaining have a less certain relationship (`relatesToXXX` predicate). Across all mappings, 60% of all UMBEL reference concepts (or 16,884) are now linked directly to Wikipedia via the new `umbel:correspondsTo` property. Across all of these mappings, nearly 4 million predicate relations (3,935,148) link UMBEL to Wikipedia.

Three methods have been employed to link Wikipedia pages (instances) via the DBpedia (v. 3.51) extraction to the UMBEL reference concept structure:

- In method one, the instances associated with the DBpedia ontology have been inherited directly based on their class mappings to the UMBEL reference concepts. These mappings also receive the `rdf:type` predicate. Some 659,527 unique pages have been linked in this matter, resulting in a total of 876,125 `rdf:type` assignments;
- In method two, using Semantic Vectors applied to "clean" DBpedia categories, an association file to candidate UMBEL reference concepts with SV scores has been created for every clean DBpedia category. These candidates have been then inspected by hand with an assignment made manually. DBpedia instances associated with these categories have been then mapped to the UMBEL structure and given a `relatesToXXX` predicate for the reference concept's associated, single ST (SuperType). Because multiple DBpedia instances could be related to different reference concepts, then individual DBpedia pages may have been assigned multiple `relatesToXXX` predicates. (If the DBpedia page already had a `rdf:type` assignment, this would supercede the `relatesToXXX` predicates);
- With method two, 2,484 unique reference concepts participated in the linkage to 102,956 unique DBpedia pages. A total of 111,470 `relatesToXXX` predicates have been created based on this method;
- In method three, the Wikipedia categories have been deconstructed to discern their structural compositions, largely based on suffix extensions. A script has been used to relate these DBpedia categories by list to candidate UMBEL reference concepts. These lists have been then presented via script for assigning by hand to the associated UMBEL reference concept. Instances related to the assigned DBpedia category have been then given the same `relatesToXXX` predicate that was associated with the related UMBEL reference concept;
- With method three, 1,668 unique reference concepts particated in the linkage to 1,808,782 unique DBpedia pages. A total of 2,947,553 `relatesToXXX` predicates have been created based on this method;
- Lastly, a fourth source, not really a method, adds 7,405 Wikipedia instances by virtue of hand-inspected OpenCyc [9] to DBpedia page mappings within the current OpenCyc knowledge base.

Two ancillary objectives have been included in the effort to secure UMBEL's role as a central reference ontology. These two objectives are: 1) to add further reference concepts (RCs) to UMBEL's core ontology, and 2) to refine UMBEL's existing vocabulary with additional linking predicates (`relatesToXXX` predicates).

Three major changes to the UMBEL vocabulary and reference concept structure (ontology) have been made as the result of this effort:

(a) The first major change has been to add 7,405 reference concepts to the core UMBEL structure. These additions came about as a way to complete the coverage of the general UMBEL structure in order to provide appropriate linkage points into the ontology. The analysis leading to these additions came about from analyzing existing OpenCyc to DBpedia linkages and missing linking concepts due to the DBpedia and GeoNames class mapping activities. This larger "core" UMBEL structure is now felt to be closer to adequate for ongoing reference mappings to other external ontologies into the future.

(b) The second major change has been to add 31 new predicates to the UMBEL vocabulary to represent a linkage relationship to a SuperType. These predicates all have the form `relatesToXXX`, for instance `relatesToAbstraction`, `relatesToActivity`, `relatesToAnimal`, etc. The predicate indicates that the object instance has a relation to the SuperType, perhaps as a true class member or perhaps only as an attribute, but that the degree of this relationship cannot be resolved. These predicates and their association with SuperTypes is shown in 0.

(c) The third major change was to apply the UMBEL `hasMapping` predicate to all of the possible assignments, using a controlled vocabulary for characterizing the mapping assignment.

The consistency of the mappings is checked with the help of ontology building tools, like TopBraidComposer in the case of PROTON and Protege or PELLET reasoner in the case of UMBEL.

These mappings give the opportunity to use either PROTON or UMBEL to make queries over the large amount of interconnected data from different datasets. The interoperability between the datasets increases when the intermediary layers are being introduced because of several reasons. The intermediary layers ensure consistency in the interpretation of the concepts referred to in their ontologies. The intermediary layers ensure unified access over large amount of heterogeneous data. The intermediary layers ensure easier management of reason-able views of LOD by facilitating their extendability.

The approach of matching UMBEL to DBpedia is semi-automated and presents a method which can be applied to the matching of new ontologies into the reason-able view.

Thus, the KRI reason-able view includes upper-level ontologies mapped to DBPedia and a unified access point, allowing to extend the reason-able view with additional datasets and in the same time making use of the same access point.

The experimental reason-able view for KRI contains **PROTON** and its mappings to DBPedia. As PROTON is used to bridge the gap between the GF representations and SPARQL queries (cf. deliverable 4.1 [30]), it will serve as a bridge (access point) to the entire knowledge base incorporated in the reason-able view.

DBpedia, **Geonames**, **WordNet** and **PROTON** give a representative collection of datasets to be integrated as a KRI reason-able view, which allows experimentation with the KRI use cases.

This basic reason-able view will be extended in two ways:

- Adding knowledge to allow for multilingual support

- Adding domain specific schema and instance data for domain specific exploration

The first will integrate a bilingual WordNet into the reason-able view. The second will integrate a cultural heritage ontology, CRM (Core reference model) [4], and instance data from the Gothenburg city museum [19].

They are described in the next two sections.

6 MUSEUM ONTOLOGY

The CIDOC Conceptual Reference Model (CRM) is an object oriented ontology developed by the International Council of Museum's Committee for Documentation (ICOM-CIDOC) [4]; an international organization of museums and museum professionals that is devoted to conserve and communicate to society the world's nature and cultural heritage. The ontology was developed between 1994 and 2005 and it has been accepted as International Standard (ISO 21127) in September 2006.

The overall scope of CRM ontology is curated knowledge of museums. This includes all information necessary for the documentation of cultural heritage collections at the quality of serious academic work but also understandable for the large public providing the level of detail and precision necessary to museum professionals to perform their work well. The CRM has to cover all sorts of heritage objects, e.g. material collected and displayed by museums and related institutions, sites, monuments of natural history, ethnography, archeology, historic monuments and fine and applied art collections, enabling exchange of relevant information with libraries and archives. CRM is specifically intended to cover contextual information: the historical, geographical and theoretical background in which individual items are placed and which gives them much of their significance and value. Technologically, CRM aims to leverage contemporary technology while enabling communication with legacy systems.

The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modeling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives [4]. The model also provides a transportation format for data integration and thereby serves various needs like exchange of data on-line navigation.

The CRM ontology is derived from the underlying semantics of database schemata and document structures used in cultural heritage and museum documentation. It provides a common frame of reference for all organizations and thereby enables semantic interoperability.

The CRM aims to support the following specific functionalities [5]:

- Inform developers of information systems as a guide to good practice in conceptual modeling, in order to effectively structure and relate information assets of cultural documentation.
- Serve as a common language for domain experts and IT developers to formulate requirements and to agree on system functionalities with respect to the correct handling of cultural contents.
- Serve as a formal language for the identification of common information contents in different data formats; in particular to support the implementation of automatic data transformation algorithms from local to global data structures without loss of meaning. The latter being useful for data exchange, data migration from legacy systems, data information integration and mediation of heterogeneous sources.
- To support associative queries against integrated resources by providing a global model of the basic classes and their associations to formulate such queries.
- It is further believed, that advanced natural language algorithms and case-specific heuristics can take significant advantage of the CRM to resolve free text information into a formal logical form, if that is regarded beneficial. The CRM is however not thought to be a means to replace scholarly text, rich in meaning, by logical forms, but only a means to identify related data.

Users intending to take advantage of the semantic interoperability offered by the CIDOC-CRM need to annotate their data in a way that is compatible with this formal ontology. Compatibility may pertain either to the associations by which users would like their data to be accessible in an integrated environment, or to the contents intended for transport to other environments, allowing encoded meaning to be preserved in a target system.

CRM ontology consists of about 90 classes and 148 properties. The overall ontology represents an upper-level ontology view for cultural and natural history. Its higher level concepts are upper-level general concepts, e.g. `Entity`, `Temporal Entity`, `Time Span`, `Place`, `Dimension`, `Persistent Item`. The `Persistent Item` class comprises items that have a persistent identity, sometimes known as "endurants" in philosophy. They can be repeatedly recognized within the duration of their existence by identity criteria rather than by continuity or observation. `Persistent Items` can be either physical entities, such as people, animals or things, or conceptual entities such as ideas, concepts, products of the imagination or common names.

The remaining ontology concepts are described under two of these upper-level concepts, namely `Temporal Entity and Persistent Item`. Still they appear at the third and fourth level of depth, preceded by another set of upper-level concepts such as `Actor` (`Person` and `Group`), `Thing` (`Man Made Thing` and `Legal Object`).

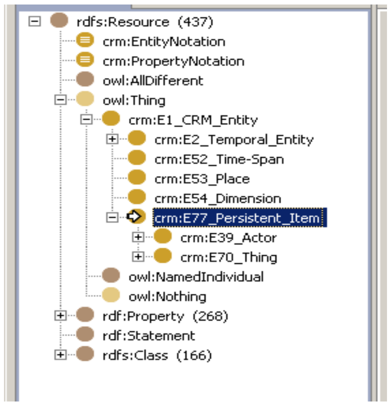

Figure 5

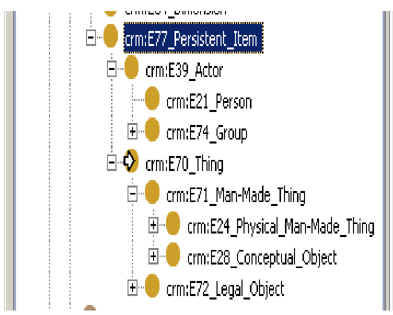

Figure 6

Such museum related concepts are described as `Man-made-thing`, and `Conceptual Object` and refer to `Visual Items`, `Linguistic Objects`, `Information Objects`, all of them are concepts of identifiable items some of them inheriting from several super concepts, cf. figure 7.

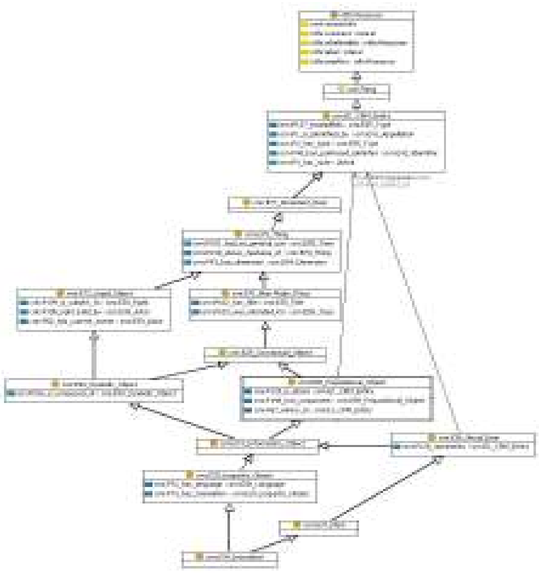

Figure 7

It is important to emphasize that the CRM model is conceived in a way, that it separates the physical objects from their actual content. For example, the concept `Visual item` is defined to refer to the content of the visualization, i.e. a content bearing object, a logo or a portrait, excluding the actual physical support it is produced or printed on.

Another area in the CRM ontology where the museum and cultural heritage information is encoded is the specialization of the concept Event, which includes as sub concepts the entire lifecycle of an artefact, e.g. Production, Creation, Dissolution, Acquisition, Curation, etc. as shown on figure 8.

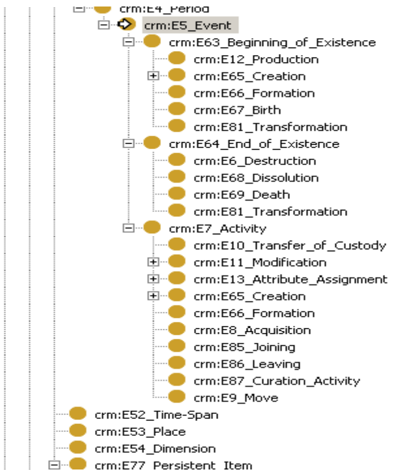

Figure 8

The CRM ontology is designed as an object-oriented semantic model. The idea behind this design is to make it understandable to both documentation experts and information scientists, while at the same time being readily convertible to machine-readable formats such as RDF Schema, OWL, etc. It can be implemented in any Relational or object-oriented schema. CRM instances can also be encoded in RDF, XML, OWL and others.

The integration of CRM into KRI reason-able view takes place at the schema level by providing mappings between the CRM concepts and PROTON concepts.

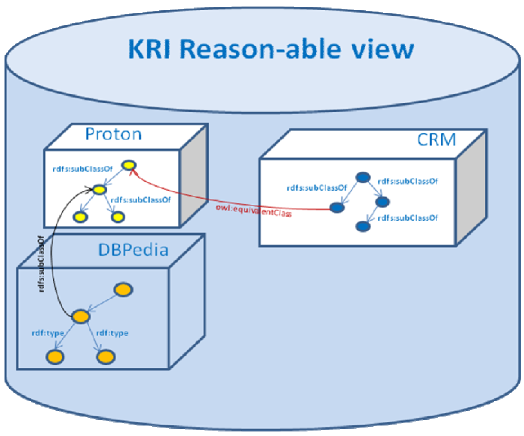

Figure 9

The CRM concepts were linked to PROTON concepts with the property `owl:equivalentClass`. This reflects the information of the two classes. As PROTON is a very concise upper-level ontology, only a couple of CRM concepts were linked to PROTON's concepts, e.g. `crm:E5_Event`, `crm:E52_Time_Span`, `crm:E53_Place`, `crm:E39_Actor` , `crm:E84_Information_Carrier`, `crm:E21_Person`.

7 GOTHENBURG MUSEUM DATA

K-samsök [24], the Swedish Open Cultural Heritage (SOCH), is a Web service for applications to retrieve data from cultural heritage institutions or associations with Cultural Heritage information. K-samsök is the join between the cultural heritage institutions' databases and the institutions, companies, individuals and associations who want to use the information in their own applications for example on web pages or in mobile applications. It makes the access to Swedish cultural heritage easier and allows more people to benefit from all museum information. 25 Swedish cultural heritage institutions (museums, libraries, archives) have provided their data to K-samsök that harvests them via on-line databases and datasets. K-samsök is run by National Heritage Board [31] in collaboration with several cultural institutions (the National Archives, National Library, Western Heritage, National Historical Museums, etc.) and museum systems (including Carlotta and Primus). K-samsök is managed since January 2011 by the National Heritage Board.

A uniform data representation format of cultural heritage/museum items has been developed by museums and cultural institutions in collaboration with the Swedish museum coordinator. This representation is used by K-samsök to gather and represent information about museum items from several museums in Sweden. This uniform representation which is available in RDF includes features which are divided in the following categories:

(a) Identification of the item in the collection
(b) Internet address, and thumbnail address
(c) Description of the item
(d) Description of the presentation of the item, including a thumbnail
(e) Geographic location coordinates
(f) Museum information about the item
(g) Context, when was it created, to which style it belongs, etc.
(h) Item specification, e.g. size, and type of the item – painting, sculpture and the like

An example of an item representation that is coded in RDF according to this unified representation of Swedish museum items is given in Appendix 1 (section 14.1)

The same information about the museum item is visualized on the museum website, as shown in Figure 10. Figure 10 presents a painting item from The History Museum in Sweden that is available from the following URL:

`http://mis.historiska.se/mis/sok/fid.asp?fid=96596&g=1`

Figure 10

Looking at the information represented for K-samsök, it becomes obvious that the CRM schema is not enough to cover all of it. In order to provide the necessary infrastructure to load the complete information about a museum item, it is required to integrate the schema of K-samsök into the KRI reason-able view. This is possible by defining a new intermediary layer described in a specific ontology, say Museum Artefacts Ontology (MAO). Such an ontology will include concepts reflecting the K-samsök schema and allowing to integrate the data from museums. It is important to note, that this Museum Artefacts Ontology can be further specified with descriptions of additional concepts covering a specific type of museum artefacts, like for example the painting ontology.

The Gothenburg city museum data, which will be used as experimental data for this use case for KRI is organized according to the CIDOC-CRM structure. So, in order to be able to integrate the Gothenburg city museum's data into the KRI knowledge base, it will be necessary to build the museum artefacts ontology (MAO). This ontology will cover the concepts and relationships that reflect the K-samsök categorization system and the data available in the Gothenburg city museum. Figure 11 shows how elements from the Gothenburg museum are represented with elements from different schemata, e.g. PROTON, CRM and MAO. Additionally, the linkage with external to the Gothenburg museum data, e.g. from DBpedia, is provided by connecting MAO ontology concepts to DBpedia instances, or by connecting the Gothenburg museum data with the corresponding DBpedia instances using the predicate `owl:sameAs`.

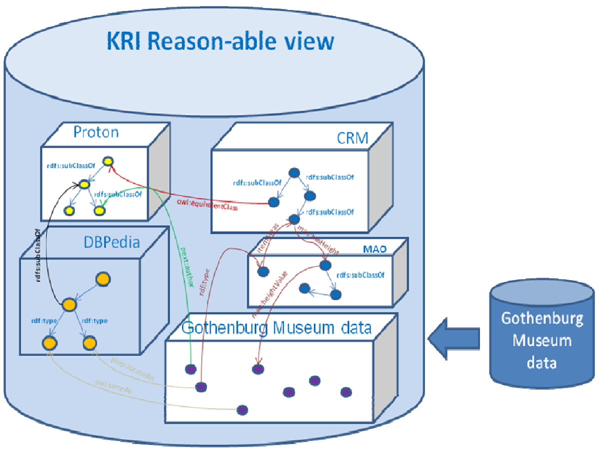

Figure 11

The Gothenburg museum preserves 8900 museum objects described in its Database. These objects correspond to two museum collections (GSM and GIM) and are placed in two tables of the museum database. 39 properties refer to each museum object, including its identification, its type - a painting, a sculpture, etc.-, its material, its measurements, its location, its inventarization. All properties are described in Swedish.

The process of Gothenburg museum data integration into the KRI knowledge base consists in transforming the information from the museum database into RDF triples based on the ontologies in it. The transformation follows the schema presented in figure 12, which shows the triples generation method for 15 of the 39 database properties.

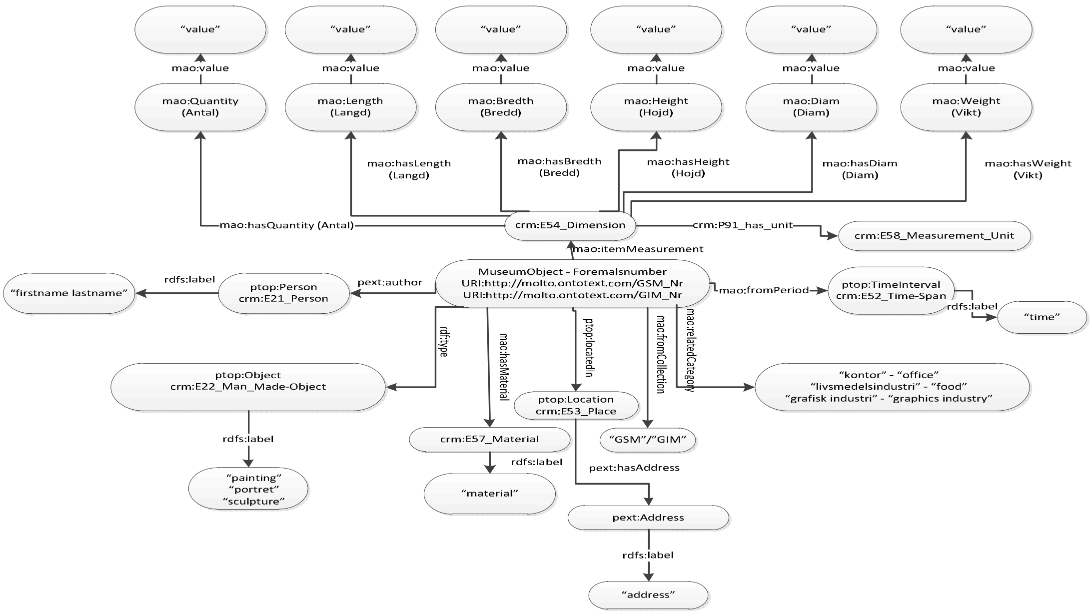

Figure 12

The triple generation goes through a process of localization, e.g. using English words for the naming of the properties and URIs in the KRI knowledge base.

Loading the KRI knowledge base with the Gothenburg museum data allows to formulate queries of the following nature:

- Museum artefacts preserved in the museum since 2005
- Paintings from the GSM collection
- Inventory numbers of the paintings from the GSM collection
- Location of the objects created by Anders Hafrin
- Paintings with length less than 1 m
- etc.

Having this infrastructure and knowledge base in place built according to the models described above will provide an easy path to extension of the knowledge base with data from other Swedish museums or generally museum data, giving the opportunity to query and obtain results about artefacts belonging to different museum collections.

Figure 13 below shows the results from the SPARQL query "Museum Objects from Swedish Museums"

```
select ?museumObject ?museum where {
 ?museumObject core:P109_has_current_or_former_curator ?museum .
 ?museum ptop:locatedIn dbpedia:Sweden }
```

run on the SPRAQL endpoint exposing KRI knowledge base with integrated museum data described in the sections above, e.g. http://museum.ontotext.com/sparql. Note the City museum of Gothenburg, the answer in the second column, is represented with its DBpedia URI, thus providing straightforward linking to all related information available about it in the LOD cloud.

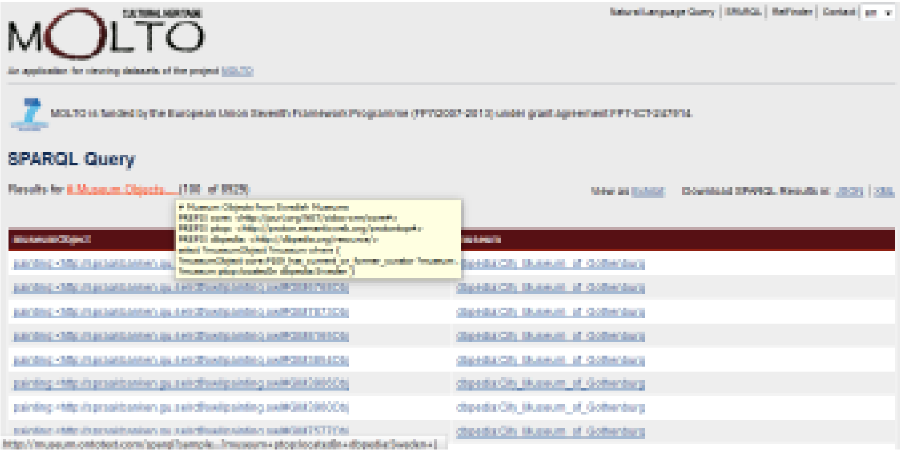

Figure 13

8 MULTILINGUAL SUPPORT

The multilingual support of KRI reason-able view is exemplified by integrating of the bilingual English-Finnish WordNet into it. The English-Finnish WordNet is developed by the MOLTO consortium partner UHEL from Finland. Before describing the characteristics of the bilingual WordNet and the adopted integration methods, we will dedicate the next paragraphs to WordNet itself.

WordNet is a very big and popular lexical knowledge base, which was designed in the 70s and 80s of the 20th century on the basis of psycholinguistic theories of human lexical memory, as an on-line lexical reference system. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. The synonym sets are linked to each other with different relations. WordNet originally contains approximately 95,600 different word forms (51,500 simple words and 44,100 collocations) organized into some 70,100 word meanings, or sets of synonyms. Following psycholinguistic analysis, the four language categories described in WordNet are organized according to different principles, reflecting a different way of categorizing experience. Nouns are organized in lexical memory as topical hierarchies, verbs are organized by a variety of entailment relations, and adjectives and adverbs are organized as N-dimensional hyperspaces. WordNet is organized around word meanings, not word forms, recognizing that a word is a conventional association between a lexicalized concept and an utterance that plays a syntactic role. The basic conceptual construct of WordNet is the Lexical Matrix, c.f. Table 1, which has word forms as headings of columns and word meanings as headings of rows.

Word Meanings	Word Forms				
	F1	F2	F3	…	Fn
M1	E1,1	E1,2			
M2		E2,2			
M3			E3,3		
.				.	
.				.	
.				.	
Mm					Em,n

Table 1

The entries in the cells say that a given word can be used to express the meaning of this row. If there are two entries in the same column, this means that the word is polysemous, if there are two entries in the same row, this means that the words are synonyms. The mappings between the forms and the meanings are many to many, e.g. some forms have several different meanings, and some meanings can be expressed by several different forms. WordNet includes descriptions of semantic relations and lexical relations. The word meanings are represented in WordNet based on the principle of differentiation, e.g. the human ability to recognize that two things are different [28], [29]. The synonym sets (synsets) of WordNet are designed to meet the differentiation principle. To make the distinction between a river bank and a bank as an institution, it is sufficient to include the word bank in two synonym sets, e.g. {bank, shore, ...} and {bank, bursary, ...}. WordNet synsets signify that a concept exists without explaining what it is. Hence the lexical matrix can be represented as a link between words and synsets. And semantic relations are designed to be pointers between synsets. The following semantic relations are defined in WordNet: (a) synonymy, (b) antonymy, (c) hyponymy, (d) meronymy.

(a) synonymy, two language expressions are synonymous if the substitution of one for the other never changes the truth value of a sentence in which the substitution is made

(b) antonymy, is a lexical relation between word forms, not a semantic relation between word meanings, as it is not always the case that the antonym of a word x actually describes not-x

(c) hyponymy is a semantic relation between word meanings. The hyponym synset is recognized if the concept it describes can be asserted as "a kind of" of a concept of another synset. Hyponymy is and asymmetrical [27], and, since there is normally a single superordinate, it generates a hierarchical semantic structure

(d) meronymy is the part-of relation. concept represented by one synset is a meronym of a concept represented another synset if native speakers of a language accept sentences constructed from such frames as A y has an x (as a part) or An x is a part of y. The meronymic relation transitive (with qualifications) and asymmetrical [8], and can be used to construct a part hierarchy (with some reservations, since a meronym can have many holonyms).

Built originally for English, WordNet was extended with versions for more than 20 languages since the 90ies. Projects like EuroWordNet and BalkaWordNet ensured the coverage of the Western European languages and Bulgarian, Czech, Greek, Romanian, Turkish and Serbian. Standalone versions of WordNet are the Russian WordNet, French and Italian WordNet which were built in separate projects. An initiative of Max Plank Institute creates a Universal WordNet covering 1.5 million words in more than 200 languages leading to a lexical knowledge base of more than 15 million words in different languages.

An RDF version of WordNet was first published in 2006. The current RDF version of WordNet is for WordNet 3.0.

The following sequences of triples show how WordNet entities are represented in RDF.

```
wn30:synset-zygotic-adjective-1 a wn30:AdjectiveSynset .
wn30:synset-zygotic-adjective-1 wn30:hasId 302882275 .
wn30:synset-zygotic-adjective-1 rdfs:label "zygotic"@en-us .

wn30:synset-president-noun-3 a wn30:NounSynset .
wn30:synset-president-noun-3 wn30:hasId 110467179 .
wn30:synset-president-noun-3 rdfs:label "president"@en-us .
```

Each word is given a URI with the following structure:

```
wn30:synset-WORD-PartOfSpeech-Number
```

This synset is assigned a category of synset, e.g. AdjectiveSynset, NounSynset or VerbSynset. It has a unique ID of nine digits, and a label, linking the synset to the actual word referring to it.

```
<wn20schema:NounSynset rdf:about="&wn20instances;synset-significant_digit-
noun-1"
                       rdfs:label="significant digit">
          <wn20schema:synsetId>112966777</wn20schema:synsetId>
</wn20schema:NounSynset>
```

WordNet is part of the LOD cloud and interlinked with DBpedia and Opencyc. The interlinking is made at the level of synsets. Thus, OpenCyc concepts are linked to WordNet synsets with the predicates `owl:sameAs` or with `rdf:type`, which means that WordNet synsets are interpreted as describing ontological concepts, e.g.

```
<http://sw.cyc.com/concept/Mx4rxrDBKCUYT1SIlLJWiUtBRA>
owl:sameAs
<http://www.w3.org/2006/03/wn/wn20/instances/synset-argumentation-
noun-2>
```

The linkage between DBPedia and WordNet is also performed at the level of synsets. However, the peculiar status of WordNet synsets, describing concepts by means of lists of synonyms was accounted for in DBpedia mappings by introducing a special predicate, not part of DBpedia ontology, but of DBpedia properties, e.g. `dbpedia-property:wordnet_type`, e.g.

```
<http://dbpedia.org/page/Nokia>
dbpedia-property:wordnet_type
<http://www.w3.org/2006/03/wn/wn20/instances/synset-company-noun-1>
```

It has to be pointed out that whereas the mapping with OpenCyc interlinks WordNet synsets at the level of concepts, the DBpedia mapping assigns instance data to wordnet synsets, thus making them accessible via the WordNet schema, cf. Figure 14.

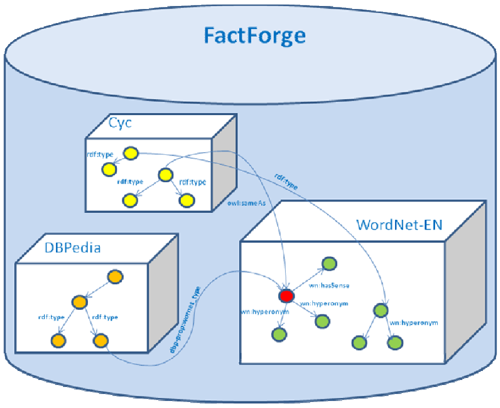

Figure 14

Verb and adjective synsets have not been linked neither with DBpedia, nor with OpenCyc. The role of verbs being rather one of properties, and the role of adjectives depending on their types being either interpretable as a property or as a special category of concept like Quality, as defined in DOLCE [13] explains this.

The objective of this section is to show how this interconnectedness will become multilingual, e.g. how will the bilingual English-Finnish WordNet be integrated into the KRI reason-able view, and what will be the impacts of this integration.

The Finnish WordNet follows the general principles of the original English WordNet, and defines wordsenses for words, identified with a unique ID, and a lexicalForm, that belong to synsets.

The word unique ID in the Finnish WordNet consists of the following components:

```
#fi-The_Word-PartOfSpeechSymbol
```

The lexical form of the word in Finnish, and the sense schema of the word are attached to the word URI, as in the English WordNet, described above.

The sense schema of a WordNet word entry consists of a WordSense and the synset the WordSense in question belongs to. The WordSense is represented as an extension of the word unique ID with indication of the WordNet version, and a sequence of 9 digits, corresponding to the word sense, e.g.

```
#fi-The_Word-PartOfSpeechSymbol-_WN30-999999999
```

The bilingual version of the Finnish WordNet connects Finnish wordnet entries with their English equivalents at the synset level. Thus, the Finnish bilingual version in OWL, introduces the predicate `sign:translationOf`. This predicate links a wordsense from the Finnish wordnet to a wordsense from the English wordnet, as shown in the example below:

```
<wn20schema:Word rdf:about="#fi-presidentti-N">
      <wn20schema:lexicalForm
xml:lang="fi">presidentti</wn20schema:lexicalForm>
      <wn20schema:sense>
         <wn20schema:WordSense rdf:about="#fi-presidentti-N_-_WN30-
110467179">
            <wn20schema:inSynset>
```

```
            <wn20schema:Synset rdf:about="&wn30;synset-president-noun-3"/>
         </wn20schema:inSynset>
         <sign:translationOf>
            <wn30:WordSense rdf:about="&wn30;wordsense-president-noun-3"/>
         </sign:translationOf>
      </wn20schema:WordSense>
   </wn20schema:sense>
</wn20schema:Word>
```

Furthermore, the wordsense in the bilingual wordnet is described as a combination between the synset of the English wordnet the word belongs to and the wordsense of the English wordnet the word is translation of. It has to be pointed out that one LexicalForm can represent different synsets or different wordsenses of the same synset, and be part of several bilingual entries.

The example below describes another intepretation of the Finnish word "`presidentti`" referring to another synset.

```
<wn20schema:Word rdf:about="#fi-presidentti-N">
      <wn20schema:lexicalForm
xml:lang="fi">presidentti</wn20schema:lexicalForm>
      <wn20schema:sense>
         <wn20schema:WordSense rdf:about="#fi-presidentti-N_-_WN30-
110467395">
            <wn20schema:inSynset>
               <wn20schema:Synset rdf:about="&wn30;synset-
President_of_the_United_States-noun-1"/>
            </wn20schema:inSynset>
            <sign:translationOf>
               <wn30:WordSense rdf:about="&wn30;wordsense-President-noun-2"/>
            </sign:translationOf>
         </wn20schema:WordSense>
      </wn20schema:sense>
   </wn20schema:Word>
```

Whereas the two examples below show two interpretations of the Finnish word "`puheenjohtaja`" sharing the same synset, but differing in the wordsenses they refer to.

```
<wn20schema:Word rdf:about="#fi-puheenjohtaja-N">
      <wn20schema:lexicalForm
xml:lang="fi">puheenjohtaja</wn20schema:lexicalForm>
      <wn20schema:sense>
         <wn20schema:WordSense rdf:about="#fi-puheenjohtaja-N_-_WN30-
110468962">
            <wn20schema:inSynset>
               <wn20schema:Synset rdf:about="&wn30;synset-president-noun-4"/>
            </wn20schema:inSynset>
            <sign:translationOf>
               <wn30:WordSense rdf:about="&wn30;wordsense-chairman-noun-1"/>
            </sign:translationOf>
         </wn20schema:WordSense>
      </wn20schema:sense>
   </wn20schema:Word>

 <wn20schema:Word rdf:about="#fi-puheenjohtaja-N">
```

```
      <wn20schema:lexicalForm
xml:lang="fi">puheenjohtaja</wn20schema:lexicalForm>
      <wn20schema:sense>
         <wn20schema:WordSense rdf:about="#fi-puheenjohtaja-N_-_WN30-
110468962">
            <wn20schema:inSynset>
               <wn20schema:Synset rdf:about="&wn30;synset-president-noun-4"/>
            </wn20schema:inSynset>
            <sign:translationOf>
               <wn30:WordSense rdf:about="&wn30;wordsense-president-noun-4"/>
            </sign:translationOf>
         </wn20schema:WordSense>
      </wn20schema:sense>
   </wn20schema:Word>
```

The schema of the interoperability between the two languages in the bilingual wordnet, and the interoperability between WordNet and DBpedia are shown on Figure 15 below.

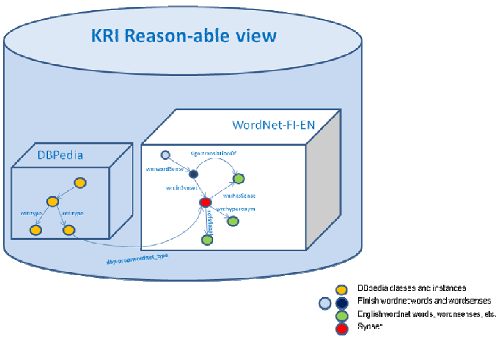

Figure 15

Having these mappings in place, the integration of the bilingual Finnish-English WordNet into the KRI reason-able view is a straightforward process. It has to be loaded into the basic KRI Reason-able view, and the interoperability is ensured.

This provides access to both languages, for instance through SPARQL [37] queries about the translation of a given Finnish word, as shown in the query:

```
select ?lexicalForm where {

?word wn20schema:lexicalForm “presidentti” .
?word wn20schema:WordSense [ sign:translationOf [ wn30:WordSense [ rdfs:label
?LexicalForm ] ] ] .

}
```

This query will return the English equivalents of all wordsenses of the Finnish word `presidentti`.

Except for queries about linguistic objects, the KRI reason-able view allows to retrieve DBpedia instances by asking with lexical forms from WordNet, both in English and in Finnish. This connectedness however requires mentioning of actual synsets or wordsenses in the query in order to obtain concrete DBpedia instances.

For example, the query using the English word “`president`” below, will retrieve results, referring to all synsets and all senses of this word.

```
select ?inst where {

?inst dbp-prop:wordnet_type ?synset .
?synset rdfs:label "president"@en-us .

}
```

To obtain a more concrete set of results, the synset has to be restricted with the mention of a wordsense.

```
select ?inst where {

?inst dbp-prop:wordnet_type ?synset .
?synset wn30:hasSense wn30:wordsense-president-noun-3 .
?synset rdfs:label "president"@en-us .

}
```

The asking about DBpedia instances with the Finnish lexical form is possible with the query:

```
select ?inst where {

?inst dbp-prop:wordnet_type ?synset .
?ws wn:inSynset ?synset .
?w wn20schema:WordSense wn30:wordsense-president-noun-3 .
?w wn20schema:lexicalForm ”presidentti” .

}
```

It also mentions a concrete wordsense explicitly.

To allow the full interoperability and more flexible querying through the bilingual wordnet without the need of knowing the exact wordnet synsets and wordsenses, however, it is necessary to provide connectedness from the wordsenses, synsets and translations to the word. This means to introduce properties which are “`inverse of`” the properties described in the schema of the bilingual WordNet, e.g.

```
sign:translationOf owl:inverseOf wn:isTranslationOf .
wn20schema:inSynset owl:inverseOf wn:synsetContainer .
wn20schema:Synset owl:inverseOf wn:synsetOf .
wn20schema:WordSense owl:inverseOf wn:wordSenseOfWord .
wn20schema:lexicalForm owl:inverseOf wn:isLexicalFormOf .
```

These properties ensure the flexibility in the usage of the predicates describing the English and Finnish equivalents. Loading the so modified bilingual Finnish-English WordNet into the KRI reason-able view provides interoperability between the two languages and DBpedia. It becomes possible to formulate queries with Finnish words, and retrieve DBpedia instances. For instance, a SPARQL query about presidents, using the Finnish word “`presidentti`” given below:

```
select ?inst where {

?inst dbp-prop:wordnet_type ?synset .
?synset wn:hasWordSense ?ws .
?ws wn:wordSenseofWord ?w .
?w wn20schema:lexicalForm "presidentti" .

}
```

will retrieve all DBpedia instances that are associated with the wordsenses and synsets of the word with lexical form "`presidentti`". Note that in this particular case there will be three synsets accounted for, e.g. `synset-president-noun-3`, `synset-chief_justice-noun`, `synset-President_of_the_United_States-noun-2`.

The retrieval results will have instances if there are assignments of them to these three synsets.

The retrieval of DBpedia instances in the described way requires not only knowing of the WordNet schema, but also knowing of the naming of the synsets and wordsenses to be used in the queries.

This is practical only in circumstances when reaching one precise instance belonging to one precise sense is important. In many cases though referring to a precise synset or to a precise wordsense would not be necessary. The following paragraphs explain an alternative way of interoperability between WordNet and DBpedia. It will also allow to obtain DBpedia instances by asking with an English or a Finnish word.

We have explained earlier in this document the approach of using intermediary layers to reach consistency and easier access to the heterogenous data in FactForge, c.f. section 5. the alternative way of interoperability between WordNet and DBpedia by mapping WordNet synsets to PROTON concepts, and using the mapping between PROTON and DBpedia schemata to reach for the instances of interest.

This approach has clear advantages:

- Gathering synsets referring to one and the same conceptual entity under a single PROTON concept
- Ensuring consistency provided by the mapping between PROTON and DBpedia at the schema-level
- Linking lexical forms and ontological concepts
- Re-using the mapping techniques for schemata designed according to different principles described in section 5
- Formulating SPARQL queries in a simpler way

We have seen that the word "`presidentti`" in Finnish has three synsets, two of which refer to a "president of a country", and one refering to a "chairman". Mapping the two synsets for president to the PROTON concept of `President`, and the one synset for chairman to the PROTON concept of `Chairman` if available ensures the linkage between the lexical forms describing the synsets and PROTON concepts. Figure 16 presents the schema of the mappings.

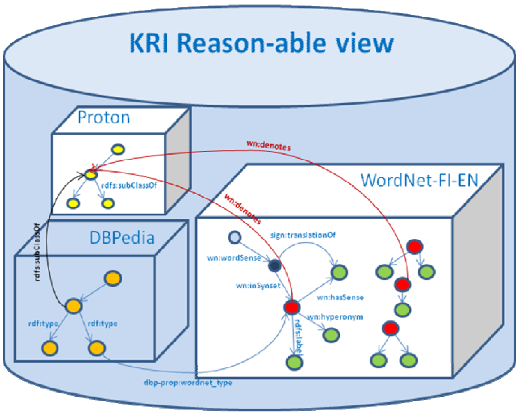

Figure 16

To preserve the conceptual difference between WordNet synsets and PROTON concepts, they are linked with the property `wn:denotes` rather than with a subsumption relation, e.g.

```
wn:synset wn:denotes proton:Concept
```

The conceptual models of WordNet and PROTON are different, as the one describes the cognitive patterns of lexicalization of human languages, and the other presents a very generic view of conceptualizing the world; therefore it is expected that not all mappings are one to one relations. Thus, the synsets describing presidents refer to persons, like in the sentences:

(i) The president is elected every four years.

(j) George Bush is the 43rd President of the United States.

In PROTON the concept of president is conceptualized as a `Title` associated with a `SocialPosition` held by a person. To solve this, the following inference rules can be applied:

```
(a)      wn:synset-president-noun-3 wn:denotes proton:President
         ----------------------------
         a proton:hasPosition ?pos
         ?pos proton:hasTitle proton:President


(b)      wn:synset-President_of_the_United_States-noun-2 wn:denotes
                                                  proton:President
         ----------------------
         a proton:hasPosition ?pos
         ?pos proton:hasTitle proton:President


(c)      wn:synset-Bush-noun-4 wn:denotes proton:President
         --------------------------------------------------
```

```
a proton:hasPosition ?pos
?pos proton:hasTitle proton:President
```

These rules reuse the method developed to cope with the mismatches in the conceptualizations between DBpedia ontology and PROTON. The rest of the mapping rules described in section 5 can be applied in suitable situations.

The rules:

```
(a)      wn:synset-president-noun-3 wn:denotes proton:President
(b)      wn:synset-President_of_the_United_States-noun-2 wn:denotes
                                                          proton:President
(c)      wn:synset-Bush-noun-4 wn:denotes proton:President
```

will allow to retrieve DBpedia instances in the following way:

(a) for English

```
select ?inst where {

?inst proton:hasPosition ?pos .|
?pos proton:hasTitle ?title .
?synset wn:denotes ?title .
?synset rdfs:label "president" .
}
```

(b) for Finnish

```
select ?inst where {

?inst proton:hasPosition ?pos .
?pos proton:hasTitle ?title .
?synset wn:denotes ?title .
?synset wn:hasWordSense ?ws .
?ws wn:wordSenseofWord ?w .
?w wn20schema:lexicalForm "presidentti" .
}
```

The returned results will be DBpedia instances of presidents.

This section described the data models and the approach of integrating WordNet with DBpedia and PROTON into a KRI reason-able view.

The integration approach of WordNet with DBpedia and PROTON will be tested in the KRI reason-able view environment to be built in the next project stage.

The next sections describe how the KRI knowledge base is used in MOTLO for building of the interface between ontologies and grammars and for the integration of domain specific data in the KRI reason-able view.

9 THE ONTOLOGY-GRAMMAR INTERFACE

The KRI ontology infrastructure allows storing, interlinking, querying, and updating ontologies used in MOLTO. In this section, we survey some uses that MOLTO makes of the infrastructure.

In connection with the museum use case, UGOT has designed a method to build text grammars for verbalization of domain ontologies. For the museum use case, the grammar consists of templates for expressing information about paintings in natural language sentences. The grammars are published and documented in [7]. A text template is in effect a vector of slots for properties expressed in the ontology, for example author, location and material. Given concrete syntaxes for the template and lexicon modules for the properties, classes and instances in the ontology, the template can be compiled into GF and used to generate multilingual verbalizations of the selected vector of properties.

We go through the verbalization of the museum ontology with regard to the different stages of natural language generation.

Content determination The content comes from the ontology. The original triples of the ontology have been translated to GF proof objects, which include `Painting`, `PaintingType`, `Painter`, `Year`, `Museum`, `Colour`, `Size` and `Material`. The proof objects guarantee that the grammar describes only existing combinations of a painting and its other properties. Content selection is done by the user, by selecting the properties to verbalize.

Document structuring The sentence structure is handled by a function that takes arguments of fixed types and inserts them into the template. In that sense, document structuring is boilerplate based. However, the boilerplates are not canned text, since they are compiled into GF grammars. The structure and grammar of the text generated can vary depending on the input and target language.

Aggregation In the abstract syntax, the type of the pattern-building function determines what kind of information is included in the pattern, and the syntactic structures are defined in the concrete syntax of each language. The concrete syntax can vary depending on how each language is able to aggregate the ontology statements. Aggregation patterns are designed by the grammar engineer of each language.

Lexical choice The words for the concepts both from the ontology (*portrait, canvas, Mona Lisa*) and additional words needed to make descriptions fluent (*display, size*) are given in the concrete syntax of each language. Some come straight from the ontology, as an object of a name or label predicate, others are decided by the grammar writer.

Referring expression generation In order not to repeat names, in the first sentence the painting is called by its name, in the second by a pronoun *it*, and in the third by its type, such as *oil painting* or *miniature portrait*. This information structure is hard-coded into the pattern; again, however, the realization is not just canned text, but the linguistic features are taken into account.

Realization Syntax, morphology and orthography are all handled by the GF resource grammar. For instance, when talking about *peinture* (feminine) and *portrait* (masculine) in French, the determiners and the modifiers are inflected accordingly.

10 ONTOLOGIES AND LEXICONS

The ontology verbalization done by UGOT is specifically designed for the art museum domain, and the proof objects are manually extracted from the Gothenburg City Museum ontology. Here we consider ways of making the process more automatic and domain independent.

An advantage of the grammar-based boilerplate model of verbalization is its simplicity. In analogy with the resource grammar API, we can envisage extending the museum case into a reusable library of verbalization/textualization patterns. In order to maintain simplicity, the GF templates can be made more abstract. Instead of hard-coding ontology specific description words (*paint, display,* size units), we generalize them as parameters chosen according to the domain and the ontology in question. UHEL has conducted some tests in this direction, generalizing the museum case patterns to more generic object description patterns.

Besides generalizing the GF text pattern grammars, the ontology-grammar mapping can be generalized using TermFactory term ontologies. TF term ontology tools allow maintaining ontology-lexicon mappings as a web editable resource in ontology format stored in KRI. Instead of having to code separately for each special domain concept its grammatical properties, ontology reasoning can be applied to infer those properties from existing entries. In order to apply existing verbalization patterns to a new domain, it then suffices to subclass the concepts of the new domain to concepts for which textual patterns already exist. In easy cases, this allows removing the grammar engineer from the domain extension workflow.

11 KRI KNOWLEDGE BASE ENVIRONMENT

The KRI knowledge base environment is built as an instance of OWLIM triples store. It provides the infrastructure to support multilingual queries based on the English-Finnish WordNet, and the knowledge to query Gothenburg museum data in a structured way. It has loaded DBPedia 3.6, PROTON, CRM and MAO ontologies, and their mappings, WordNet 3.0, the English-Finnish WordNet, the modifications to WordNet described in section 8 above, and the triplified Gothenburg museum data.

KRI knowledge base environment is built as an instance of OWLIM triple store. Its museum part, supporting WP8, provides the knowledge to query Gothenburg City Museum data in a structured way. It contains: DBPedia 3.6, Geonames, PROTON, CIDOC-CRM and MAO ontologies, and their mappings, and the triplified Gothenburg City Museum data. OWLIM performs full materialization during loading. It was expected that the available retrievable statements after loading will exceed the loaded explicit statements by about 20%. The loading statistics confirmed this expectation, e.g. the number of the loaded explicit statements was 257,774,678 triples, whereas the overall number of triples available for querying was 16% more, e.g. 305,313,536.

Experiments with the KRI knowledge base regarding querying will be carried out in the next report period of MOLTO. This will include extensions of the data models to cover detailed museum artefacts descriptions, and linkage of the multilingual support with the controlled language for querying in natural language presented in deliverable 4.1 [30].

12 CONCLUSION

This deliverable presented the methods and principles of creation of the KRI knowledge base, to be used in MOLTO use cases. It described the data models necessary to provide multilingual and domain specific support for the technology developed in the project. After introducing the notion of reason-able view of the web of data, and outlining the schema-level and instance level mapping approaches deveoped at Ontotext and used in FactForge, it showed the approach of creating a KRI reason-able view which included analysis and methods of obtaining multilingual support of the KRI by integrating a bi-lingual version of WordNet into it, and giving access to domain specific knowledge by creating the conditions to integrate Gothenburg museum data into KRI with the use of CRM ontology and the creation of an intermediary layer, MAO, ensuring full integration of the museum data. It also showed the possibility of connecting the domain specific knowledge with external knowledge from DBpedia, and of providing access to DBpedia instances via linguistic queries using WordNet.

This deliverable presented the process of triplification of data from a relational database to be integrated into the KRI of MOLTO.

The next stages of the projects foresee the actual building of the KRI knowledge base and experimenting with it from the perspective of knowledge base size and extendability and from the perspective of querying compound datasets with compound schemata.

13 REFERENCES

[1] Berners-Lee, T. (2006): Design Issues: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html.

[2] Bishop, B.; Kiryakov, A.; Ognyanoff, D.; Peikov, I.; Tashev, Z.; Velkov, R. (in press) OWLIM: A family of scalable semantic repositories. Semantic Web Journal, Special Issue: Real-World Applications of OWL. http://www.semantic-web-journal.net/content/new-submission-owlim-family-scalable-semantic-repositories.

[3] Caracciolo, C., Euzenat, J., Hollink, L., Ichise, R., Isaac A., Malaisé V., Meilicke, C., Pane, J., Shvaiko, P., Stuckenschmidt, H., Šváb-Zamazal, O., and Sváte V. Results of the Ontology Alignment Evaluation Initiative 2008, (2008).

[4] CIDOC CRM webpage http://www.cidoc-crm.org/.

[5] CRM Documentation http://erlangen-crm.org/docs/documentation_crm_owl-dl_4.2.4.pdf.

[6] Damova, M., Kiryakov, A., Simov, K., and Petrov, S. (2010) Mapping the central LOD ontologies to PROTON upper-level ontology. In Euzenat, J., Shvaiko, P. (Eds) Proceedings of Ontology Mapping Workshop at ISWC 2010, Shanghai, China. http://om2010.ontologymatching.org.

[7] Dannélls, D., Ranta, A. and Enache, R. (2012). Multilingual grammar for museum object descriptions. Project deliverable, University of Gothenburg. http://www.molto-project.eu/sites/default/files/WP8-D2_3.pdf.

[8] Cruse, D. A. 1986. *Lexical Semantics*. New York: Cambridge University Press.

[9] Cycorp. OpenCyc. http://www.cyc.com/cyc/opencyc.

[10] Damova, M., Kiryakov, A. Grinberg, M., Bergman, M., Giasson, F., Simov, K. Ontology Creation based on LOD Integration In: Semi-Automatic Ontology Development: Processes and Resources, IGI Global, Hershey PA, USA, 2011. (to appear)

[11] DBPedia, structured information from Wikipedia, http://dbpedia.org.

[12] Doerr, M., Iorizzo, D. The Dream of a Global Knowledge Network – A New Approach, ACM Journal for Computing and Cultural Heritage, Vol. 1, No. 1, Article 5, June 2008.

[13] DOLCE. http://www.loa-cnr.it/DOLCE.html.

[14] El-Mekawy, M., Östman, A. Semantic Mapping: an Ontology Engineering Method for Integrating Building Models in IFC and CITYGML. In Proceedings of the 3rd ISDE Digital Earth Summit 12-14 June, 2010, Nessebar, Bulgaria, (2010).

[15] Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Šváb-Zamazal, O., Svátek, V., Trojahn dos Santos, C., Vouros, G., Wang, S. (2009) Results of the Ontology Alignment Evaluation Initiative 2009. In Proceedings of the 4th Ontology Matching Workshop at ISWC'2009, Washington, DC., USA.

[16] FactForge, a Reason-able View to the Web of Data. http://factforge.net, http://www.ontotext.com/factforge.

[17] Fellbaum, C. ed. WordNet: An Electronic Lexical Database. MIT Press, 1998. http://wordnet.princeton.edu/.

[18] Geonames, a geographical database. http://www.geonames.org.

[19] Gothenburg City Museum. http://www.stadsmuseum.goteborg.se/prod/kultur/stadsmuseum/dalis2.nsf/0/3955ae522e06a42bc12573950046ec10?OpenDocument.

[20] International Council of Museums (ICOM). http://icom.museum/.

[21] Kiryakov, A., and Momtchev, V. *Two Reason-able Views to the Web of Linked Data.* Presentation at the Semantic Technology Conference 2009, San Jose. http://www.slideshare.net/ontotext/two-reasonable-views-to-the-web-of-linked-data.

[22] Kiryakov, A., Ognyanoff, D., Velkov, R., Tashev, Z., and Peikov, I. (2009) LDSR: Materialized Reason-able View to the Web of Linked Data. In Rinke Hoekstra and Peter F. Patel-Schneider (Eds) *Proceedings of OWLED 2009*: Chantilly, USA.

[23] Kiryakov, A., Tashev, Z., Ognyanoff, D., Velkov, R., Momtchev, V., Balev, B., Peikov, I. Validation goals and metrics for the LarKC platform. FP7 – 215535 LarKC project deliverable D5.5.2. http://www.larkc.eu/deliverable/.

[24] K-samsök http://www.ksamsok.se/om-k-samsok/anslutna-institutioner/.

[25] Linking Open Data, W3C Semantic Web Education and outreach community project, http://linkeddata.org/.

[26] Linking Open Cloud Diagramme. http://lod-cloud.net/ as of September 2010.

[27] Lyons, J. 1977. *Semantics*. 2 vols. New York: Cambridge University Press.

[28] Miller, G.A., Beckwith R., Fellbaum, C., Gross, D., Miller, K.J. Introduction to WordNet: An On-line Lexical Database. In: Journal of Lexicography, No 4, Oxford University Press, pp.235-244. http://wordnetcode.princeton.edu/5papers.pdf.

[29] Miller, G.A., Beckwith R., Fellbaum, C., Gross, D., Miller, K.J. Introduction to WordNet: An On-line Lexical Database. In: Journal of Lexicography 3 (4): 235-244, 1990.

[30] Mitankin, P., Ilchev, A. D4.1. Knowledge Representation Infrastructure. MOLTO deliverable 4.1. November 2010, http://www.molto-project.eu.

[31] National Heritage Board http://www.raa.se/cms/en/our_mission.html.

[32] OWL – Web ontology language. Overview. http://www.w3.org/TR/owl-features/.

[33] OWL Horst Dialects. http://www.ontotext.com/inference/rdfs_rules_owl.html.

[34] RDF Primer. http://www.w3.org/TR/rdf-syntax/.

[35] Semantic Vectors. http://code.google.com/p/semanticvectors/.

[36] P. Shvaiko, J. Euzenat: Ten Challenges for Ontology Matching. In Proceedings of ODBASE, 2008.

[37] SPARQL. The Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/.

[38] P. Shvaiko, J. Euzenat: A Survey of Schema-based Matching Approaches. Journal on Data Semantics, 2005.

[39] Terziev, I., Kiryakov, A., and Manov, D. (2005) *D.1.8.1* Base upper-level ontology (BULO) Guidance. Deliverable of EU-IST Project IST – 2003 – 506826 SEKT.

[40] Upper Mapping and Binding Exchange Layer (UMBEL), http://www.umbel.org/.

[41] WordNet, a lexical database for English. http://wordnet.princeton.edu/.

14 APPENDIX

14.1 UNIFORM STANDARD FOR REPRESENTING MUSEUM ITEMS IN SWEDEN


```
<rdf:RDF>
       -
       <Entity rdf:about="http://kulturarvsdata.se/shm/object/96596">
              <ksamsokVersion>0.99</ksamsokVersion>
              <buildDate>2011-02-01</buildDate>
              <collection rdf:resource="http://mis.historiska.se/rdf/collection#0"/>
              <theme rdf:resource="http://kulturarvsdata.se/resurser/theme#theme_03"/>
              <createdDate>2001-10-11</createdDate>
              <lastChangedDate>2010-03-31</lastChangedDate>
              <serviceOrganization>shm</serviceOrganization>
              <serviceName>object</serviceName>
       -
       <url>
              http://mis.historiska.se/mis/sok/fid.asp?fid=96596&g=1
       </url>
       -
       <museumdatUrl>
              http://mis.historiska.se/OAICat/shm/object/museumdat/96596
       </museumdatUrl>
       -
       <thumbnail>
              http://catview.historiska.se/catview/media/thumbnail/333070
       </thumbnail>

       <subject rdf:resource="http://kulturarvsdata.se/resurser/Subject#archaeology"/>
       <mediaType>text/plain</mediaType>
       <mediaType>image/jpeg</mediaType>
       <dataQuality rdf:resource="http://kulturarvsdata.se/resurser/DataQuality#raw"/>
       <itemType rdf:resource="http://kulturarvsdata.se/resurser/EntityType#object"/>
       <itemLicense rdf:resource="http://kulturarvsdata.se/resurser/License#pdmark"/>
       <itemLicenseUrl
       rdf:resource="http://creativecommons.org/publicdomain/mark/1.0/"/>
       <itemName rdf:nodeID="name001"/>
       <itemLabel>målning (pannå) av furu</itemLabel>
       <itemSpecification rdf:nodeID="spec001"/>
       <itemDescription rdf:nodeID="desc001"/>
       <itemMaterial rdf:nodeID="mat001"/>
       <itemNumber rdf:nodeID="num001"/>
       <itemNumber rdf:nodeID="num002"/>
       <itemMeasurement rdf:nodeID="meas001"/>
```

<itemMeasurement rdf:nodeID="meas002"/>
<itemMeasurement rdf:nodeID="meas003"/>
<context rdf:nodeID="ctx001"/>
<context rdf:nodeID="ctx002"/>
<isPartOf>http://kulturarvsdata.se/shm/inventory/3094</isPartOf>
<isPartOf>http://kulturarvsdata.se/shm/site/4175</isPartOf>
<isPartOf>http://kulturarvsdata.se/raa/bbr/21400000444123</isPartOf>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/333070</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36341</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36338</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36342</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36339</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36343</isVisualizedBy>
<isVisualizedBy>http://kulturarvsdata.se/shm/media/36340</isVisualizedBy>
-
<presentation rdf:parseType="Literal">
-
<pres:item>
<pres:version>0.99</pres:version>
<pres:entityUri>http://kulturarvsdata.se/shm/object/96596</pres:entityUri>
<pres:type>Föremål</pres:type>
<pres:id>96596</pres:id>
<pres:idLabel>SHM 3094:3A</pres:idLabel>
<pres:itemLabel>målning (pannå) av furu</pres:itemLabel>
<pres:tag>furu</pres:tag>
<pres:description>Köp 1863, medeltida målning pannå</pres:description>
-
<pres:context>
<pres:event>Tillverkad/brukad/deponerad</pres:event>
<pres:placeLabel>Sverige Gotland Väte Väte</pres:placeLabel>
<pres:timeLabel>högmedeltid (1250 e.Kr. - 1350 e.Kr.)</pres:timeLabel>
</pres:context>
-
<georss:where>
-
<gml:Point srsName="EPSG:4326">
<gml:coordinates cs="," decimal="." ts=" ">18.3642,57.4492</gml:coordinates>
</gml:Point>
</georss:where>
-
<pres:image>
-
<pres:src type="thumbnail">
http://catview.historiska.se/catview/media/thumbnail/333070
</pres:src>
-
<pres:src type="lowres">
http://catview.historiska.se/catview/media/lowres/333070
</pres:src>
-
<pres:src type="highres">
http://catview.historiska.se/catview/media/highres/333070
</pres:src>
<pres:byline>Christer Åhlin, 2009-03-06, SHMM</pres:byline>
<pres:motive>Föremål från Historiska museet</pres:motive>
<pres:copyright>SHMM</pres:copyright>
-
<pres:mediaLicense>
http://creativecommons.org/licenses/by-nc-nd/2.5/se/
</pres:mediaLicense>
<pres:mediaUri>http://kulturarvsdata.se/shm/media/333070</pres:mediaUri>
<pres:mediaUrl>http://www.historiska.se/data?bild=333070</pres:mediaUrl>

</pres:image>
-
<pres:references>
<pres:reference>http://kulturarvsdata.se/shm/site/html/4175</pres:reference>
<pres:reference>http://kulturarvsdata.se/shm/inventory/html/3094</pres:referenc
e>
-
<pres:reference>
http://kulturarvsdata.se/raa/bbr/html/21400000444123
</pres:reference>
</pres:references>
-
<pres:representations>
<pres:representation
format="HTML">http://kulturarvsdata.se/shm/object/html/96596</pres:representati
on>
<pres:representation
format="XML">http://kulturarvsdata.se/shm/object/xml/96596</pres:representation
>
<pres:representation
format="RDF">http://kulturarvsdata.se/shm/object/rdf/96596</pres:representation
>
-
<pres:representation format="Museumdat">
http://kulturarvsdata.se/shm/object/museumdat/96596
</pres:representation>
</pres:representations>
<pres:organization>Statens historiska museum</pres:organization>
<pres:organizationShort>SHM</pres:organizationShort>
<pres:service>object</pres:service>
<pres:dataQuality>Rådata</pres:dataQuality>
<pres:buildDate>2011-02-01</pres:buildDate>
</pres:item>
</presentation>
</Entity>
-
<Context rdf:nodeID="ctx001">
<contextType rdf:resource="http://kulturarvsdata.se/resurser/ContextType#use"/>
<contextLabel>Tillverkad/brukad/deponerad</contextLabel>
<fromTime>1250</fromTime>
<toTime>1349</toTime>
<fromPeriodName>högmedeltid</fromPeriodName>
<toPeriodName>högmedeltid</toPeriodName>
<fromPeriodId>420</fromPeriodId>
<toPeriodId>420</toPeriodId>
<periodAuth>http://mis.historiska.se/rdf/period</periodAuth>
<country rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/country#SE"/>
<geoF:county
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/county#9"/>
<geoF:province
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/province#Go"/>
<geoF:municipality
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/municipality#980"/>
<geoF:parish
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/parish#981"/>
<countryName>Sverige</countryName>
<countyName>Gotland</countyName>
<provinceName>Gotland</provinceName>
<municipalityName>Gotland</municipalityName>
<parishName>Väte</parishName>
<placeName>Väte</placeName>
<placeName/>
-

<coordinates rdf:parseType="Literal">
-
<gml:Point srsName="EPSG:4326">
<gml:coordinates cs="," decimal="." ts="
">18.3642,57.4492</gml:coordinates>
</gml:Point>
</coordinates>
</Context>
-
<Context rdf:nodeID="ctx002">
<contextType
rdf:resource="http://kulturarvsdata.se/resurser/ContextType#explore"/>
<contextLabel>Undersökt</contextLabel>
<fromTime/>
<toTime/>
<country rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/country#SE"/>
<geoF:county
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/county#9"/>
<geoF:province
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/province#Go"/>
<geoF:municipality

rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/municipality#980"/>
<geoF:parish
rdf:resource="http://kulturarvsdata.se/resurser/aukt/geo/parish#981"/>
<countryName>Sverige</countryName>
<countyName>Gotland</countyName>
<provinceName>Gotland</provinceName>
<municipalityName>Gotland</municipalityName>
<parishName>Väte</parishName>
<placeName>Väte</placeName>
<placeName/>
-
<coordinates rdf:parseType="Literal">
-
<gml:Point srsName="EPSG:4326">
<gml:coordinates cs="," decimal="." ts=" ">18.3642,57.4492</gml:coordinates>
</gml:Point>
</coordinates>
</Context>
-
<ItemNumber rdf:nodeID="num001">
<type>Inventarienummer</type>
<number>3094</number>
</ItemNumber>
-
<ItemNumber rdf:nodeID="num002">
<type>Undernummer</type>
<number>3A</number>
</ItemNumber>
-
<ItemName rdf:nodeID="name001">
<type>Sakord</type>
<name>målning</name>
</ItemName>
-
<ItemSpecification rdf:nodeID="spec001">
<type>Typ</type>
<spec>pannå</spec>
</ItemSpecification>
-
<ItemMaterial rdf:nodeID="mat001">
<type>Material</type>

```
      <material>furu</material>
      </ItemMaterial>
      -
      <ItemDescription rdf:nodeID="desc001">
      <type>Beskrivning</type>
      <desc>Köp 1863, medeltida målning pannå</desc>
      </ItemDescription>
      -
      <ItemMeasurement rdf:nodeID="meas001">
      <type>Antal</type>
      <value>1</value>
      <unit/>
      <qualifier/>
      </ItemMeasurement>
      -
      <ItemMeasurement rdf:nodeID="meas002">
      <type>Bredd</type>
      <value>295</value>
      <unit>mm</unit>
      <qualifier/>
      </ItemMeasurement>
      -
      <ItemMeasurement rdf:nodeID="meas003">
      <type>Längd</type>
      <value>615</value>
      <unit>mm</unit>
      <qualifier/>
      </ItemMeasurement>
</rdf:RDF>
```