D1.1.2 Semantic Data Layer First Operational Version

This deliverable is software.

ECP-2008-DILI-528001

EuropeanaConnect

Semantic Data Layer First Operational Version

Deliverable number/name	*D1.1.2*
Dissemination level	*Public*
Delivery date	*16 August 2010*
Status	*v.1.0*
Author(s)	*Victor de Boer (VUA)* *Antoine Isaac (VUA)* *Jacco van Ossenbruggen (VUA)* *Guus Schreiber (VUA)* *Jan Wielemaker (VUA)* *Steffen Hennicke (HUB)*

Distribution

Version	Date of sending	Name	Role in project
0.1	05.08.2010	Luca Dini, Ross King	CELI, ARC/AIT
1.0	06.08.2010	Antoine Isaac	Researcher
1.1	16.08.2010	Antoine Isaac	Researcher

Approval

Version	Date of approval	Name	Role in project
1.1	06.08.2010	Ross King	ARC/AIT
1.1	16.08.2010	Luca Dini	Task Leader
1.1	16.08.2010	Sigfrid Lundberg	KB-DK
1.1	18.08.2010	Max Kaiser	PC

Revisions

Version	Status	Author	Date	Changes
0.1	draft	Antoine Isaac	04.08.2010	Initial version
1.0	draft	Antoine Isaac	06.08.2010	Taking into account review comments from RK
1.1	draft	Antoine Isaac	16.08.2010	Taking into account review comments by LD, SL

Table of Contents

Description of the Semantic Layer

Objective and nature of content

The Semantic Layer built by EuropeanaConnect WP1.1 aims at providing a uniform, machine-actionable, web-enabled access to the reference knowledge capitalized by the various stakeholders of Europeana.eu. This knowledge mostly comes in the form of controlled vocabularies: thesauri, subject heading lists, classification schemes, authority list for person names and place names, etc. Those are consistently used in the metadata describing the objects ingested in Europeana.eu. The objective is to make this knowledge available for enabling the Europeana.eu users to benefit from semantics-intensive functions, as will be specified in other WP1 deliverables.

The Semantic Layer thus primarily consists of data and not software. The web applications mentioned in this report are just for testing and "getting the flavour" of the data, not for production use.

To match the aforementioned objectives, the controlled vocabularies that form the Semantic Layer have been converted to the RDF format, using the SKOS model.[1] This allows a uniform representation of the concepts present in the vocabulary; it also paves the way for semantically aligning those concepts, as will be done in WP1.2.

As RDF resources, the elements of converted vocabularies (in SKOS terms, "concepts") are provided with URI identifiers. The main elements of SKOS used to describe these concepts are:

- labels, either preferred, alternative or hidden;
- semantic relations with other concepts, e.g. "broader" or "related";
- documentation notes, such as definition, scope notes.

An example of such concepts is shown in Fig. 3.

Note on difference between D1.1.1 and D1.1.2

This deliverable is an updated version of D1.1.1 "Semantic Data Layer First Prototype". The only difference lies in the amount of data that is present in the Semantic Layer, as required by the Description of Work.

Access to data

The user will in the very near future be able to access the SKOS/RDF data from the following SVN repository: http://sandbox08.isti.cnr.it/econnwp1svn (same login and password as on the EuropeanaConnect Liferay environment).

---

[1] http://www.w3.org/2004/02/skos/. For the moment lexical resources like Wordnet are still not using the SKOS constructs. But they follow a SKOS-like, concept-based modelling approach. Only synsets are explicitly represented, and the senses appear as mere labels ("senseLabel" property) attached to these elements.

For a more human-friendly exploration, the data has been loaded in an instance of the ClioPatria[2] Semantic Search server: http://semanticweb.cs.vu.nl/europeana/session/thesaurus.[3]

From the first page there (Fig.1) user can browse the various vocabularies of the Semantic Layer, using the semantic hierarchy that connects the concepts from these vocabularies (Fig. 2). User can then access the information stored for individual concepts, as shown in Fig.3.
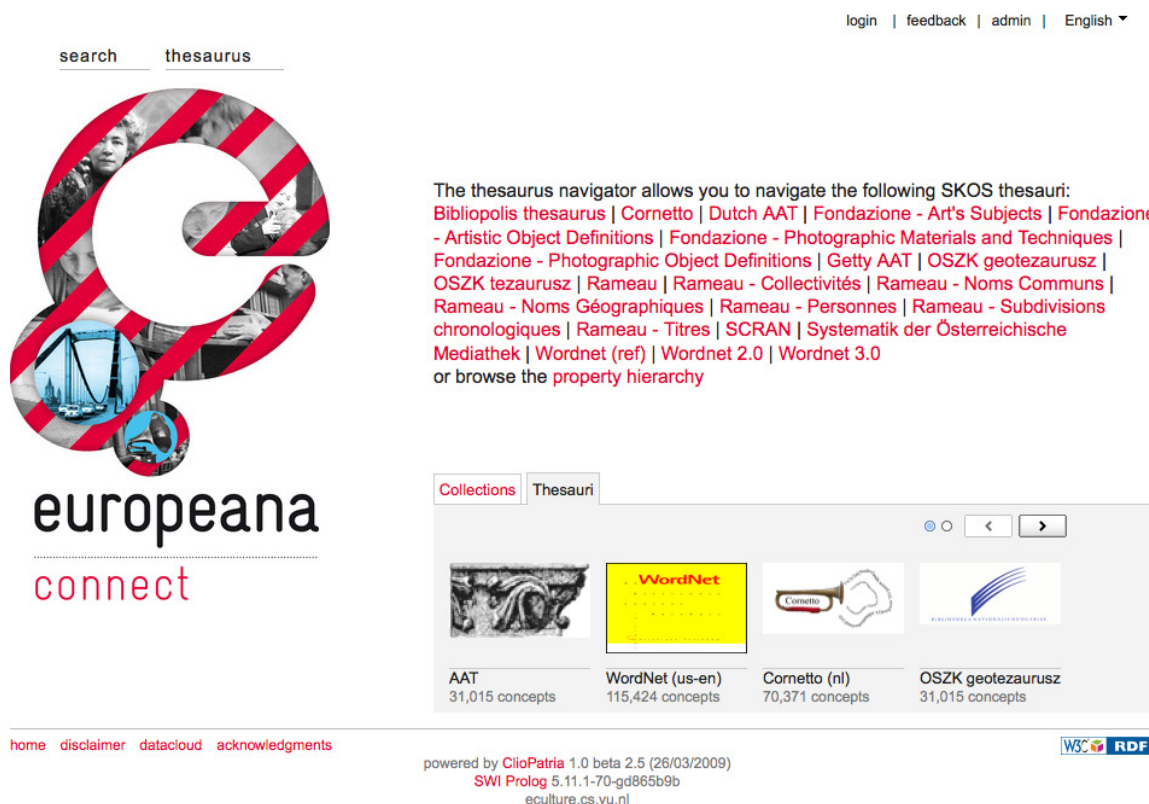

Fig. 1 Semantic Layer access interface

Fig. 2 Browsing down a vocabulary hierarchy

Fig. 3 Specific concept information

It is also possible for software agents to query the data using a SPARQL endpoint
http://semanticweb.cs.vu.nl/europeana/sparql/ that implements the SPARQL protocol
http://www.w3.org/TR/rdf-sparql-protocol/. A more human-friendly SPARQL query input interface
using the query syntax of http://www.w3.org/TR/rdf-sparql-query/ can be found at

http://semanticweb.cs.vu.nl/europeana/user/query. The reader should however be aware that this endpoint is provided *for demonstration purposes only*. It does not come with any quality assurance commitment, and no production-level systems should be built on top of it.

Current Coverage

The vocabularies and collections were obtained from a subset of the voluntary providers listed in the M1.1.1 document ("Inventory list of vocabularies finalised"). The effort of gathering and converting vocabularies and collections from that list is still ongoing. At the moment this report is submitted, the Semantic Layer contains SKOS/RDF data for 19 vocabularies. The following table gives a qualitative insight on the current content of the Semantic Layer. Statistics will be maintained and referred to from the "human-readable" page for the Semantic Layer, at http://semanticweb.cs.vu.nl/europeana/session/thesaurus. Meanwhile, readers can get a first insight on the metrics of loaded RDF files at http://semanticweb.cs.vu.nl/europeana/browse/list_graphs.[4]

Note that in the table datasets are grouped in wider "families" of vocabularies when applicable. For example the National Library of Hungary provided both a subject thesaurus and a place thesaurus, which together form a single indexing system. This is reflected by the combination of left and right alignment in the first column.

---

[4] For example,
http://semanticweb.cs.vu.nl/europeana/browse/list_resource?r=http://www.w3.org/2004/02/skos/core%23Concept indicates that at the time of writing this report, the repository contains 294,538 instances of the skos:Concept class.

Table: Semantic Layer Contents

Vocabulary	Brief description
Bibliopolis thesaurus[5]	Glossary of subjects used for print book indexing
Cornetto[6]	Semantic network of Dutch word meanings, similar to Wordnet
Dutch AAT[7]	Dutch version of Getty's Art and Architecture Thesaurus
Fondazione Zeri[8]	Vocabularies created by ICCD (Italian Ministry for Cultural Heritage) used in the catalogue of the Fondazione Federico Zeri
Art's Subjects	Subjects
Artistic Object Definitions	Materials and genres of objects
Photographic Materials and Techniques	Classification for photographic techniques
Photographic Object Definitions	Basic classification for photographs
Getty AAT[9]	Art and Architecture Thesaurus
OSZK thesauri[10]	Thesauri at the National Library of Hungary
geothesaurus	Places
thesaurus	Subjects
RAMEAU[11]	Subject Thesauri at the French National Library
collectivités	Organisations
noms communs	Common nouns
noms géographiques	Places

---

[5] http://www.bibliopolis.nl/index_en.html

[6] http://www2.let.vu.nl/oz/cornetto/

[7] http://www.aat-ned.nl/

[8] http://fe.fondazionezeri.unibo.it

[9] http://www.getty.edu/research/conducting_research/vocabularies/aat/

[10] http://www.oszk.hu/

[11] http://rameau.bnf.fr

Vocabulary	Brief description
personnes	Persons
subdivisions chronologiques	Time periods
titres	Titles of works
Systematik der Österreichischen Mediathek[12]	Thesaurus used at the Austrian media library
SCRAN classification[13]	Curriculum and topic classification of the SCRAN portal
Wordnet	Princeton's semantic network of general English word meanings
Wordnet 2.0[14]	Version 2.0, used in legacy projects such as the ThoughtLab
Wordnet 3.0[15]	Newly produced linked data version
Wordnet (ref)	A time-agnostic version of Wordnet which glues together versions 2.0 and 3.0

[12] http://www.mediathek.at/

[13] http://www.scran.ac.uk/

[14] http://www.w3.org/TR/wordnet-rdf/

[15] http://semanticweb.cs.vu.nl/lod/wn30/

Description of software developed for Europeana within EuropeanaConnect

Note that because this deliverable is primarily (RDF) data, a number of features in this table are marked NA.

Link to software	http://semanticweb.cs.vu.nl/europeana/session/thesaurus http://semanticweb.cs.vu.nl/europeana/browse/list_graphs http://sandbox08.isti.cnr.it/econnwp1svn (SVN access to RDF data)
Login information	No username or password required for web interfaces Europeana LifeRay login required for SVN access
Development environment	NA
Programming language used	NA
Application server used	NA
Database requirements	NA
Operating system requirements	NA
Port requirements / default ports used	NA
Interface	NA
Licensing conditions	Various conditions are associated with the data. Generally the data is for the sole purpose of usage within the Europeana community.