D1.1.1 Semantic Data Layer First Prototype

This deliverable is software.

co-funded by the European Union

The project is co-funded by the European Union, through the **eContent*plus*** programme

http://ec.europa.eu/econtentplus

Österreichische Nationalbibliothek

EuropeanaConnect is coordinated by the Austrian National Library

ECP-2008-DILI-528001

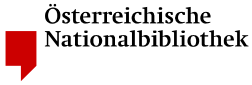
EuropeanaConnect

Semantic Data Layer First Prototype

Deliverable number/name	*D1.1.1*
Dissemination level	*Public*
Delivery date	*19 April 2010*
Status	*v.1.0*
Author(s)	*Antoine Isaac (VUA)* *Jacco van Ossenbruggen (VUA)* *Guus Schreiber (VUA)* *Jan Wielemaker (VUA)* *Steffen Hennicke (HUB)*

e*Content*plus

This project is funded under the *e*Content*plus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.



EuropeanaConnect is coordinated by the Austrian National Library

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. 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 to have 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.

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).

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.

[1] http://www.w3.org/2004/02/skos/. For the moment lexical resources like Wordnet are still not using the SKOS constructs. But they follows 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.

[2] http://e-culture.multimedian.nl/software/ClioPatria.shtml

[3] Note that the “thesaurus” section is the only one that should be explored as part of the deliverable. http://semanticweb.cs.vu.nl/europeana/session/search, especially, does not give a stable access means to the content of the semantic layer.

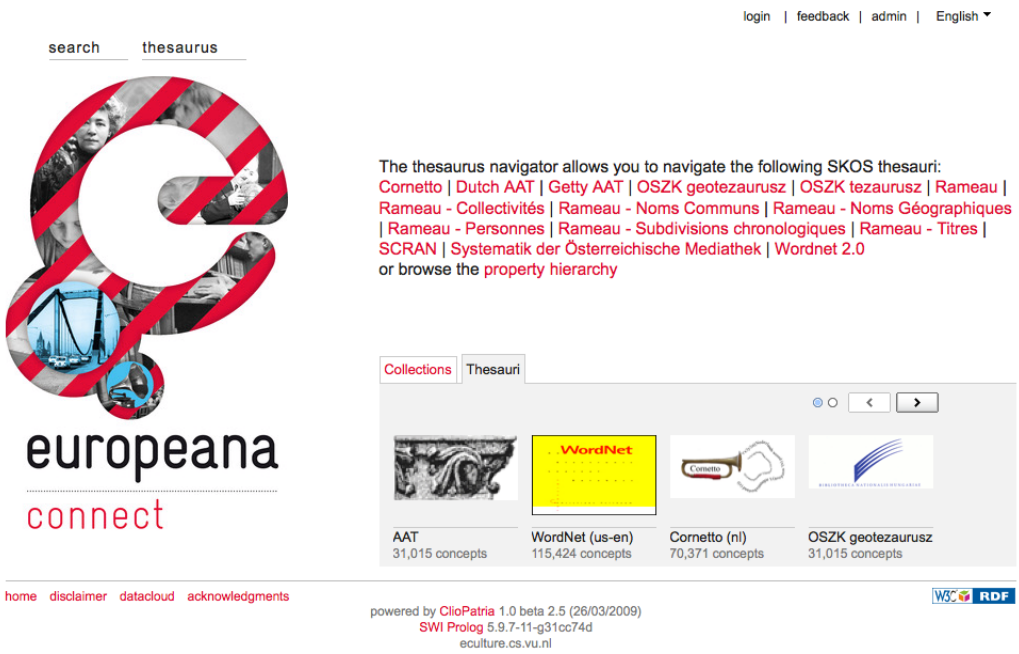


Fig. 1 Semantic Layer access interface

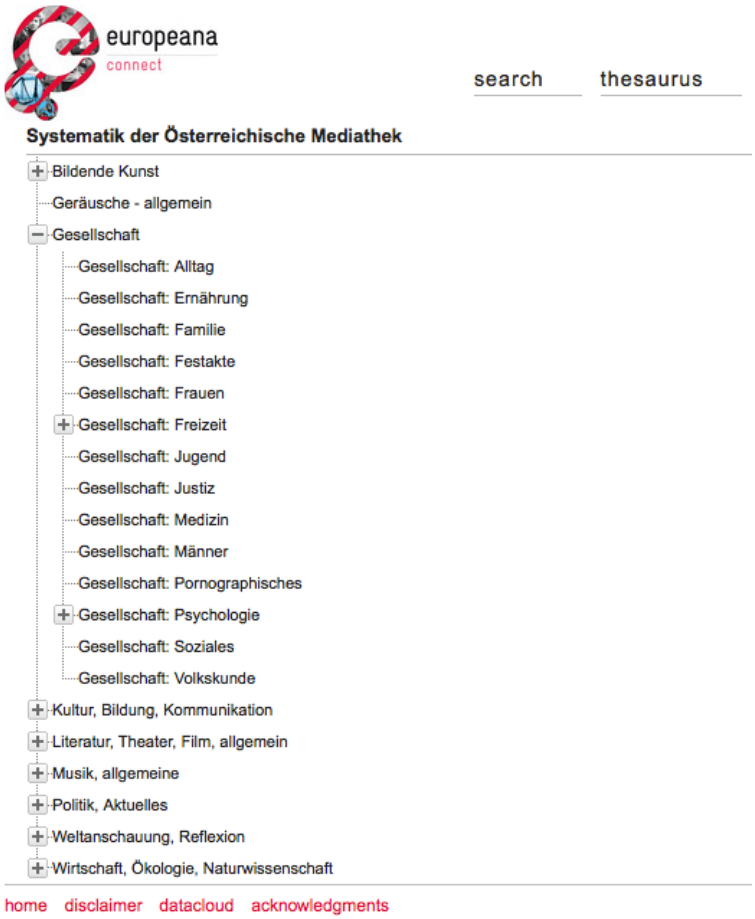


Fig. 2 Browsing down a vocabulary hierarchy

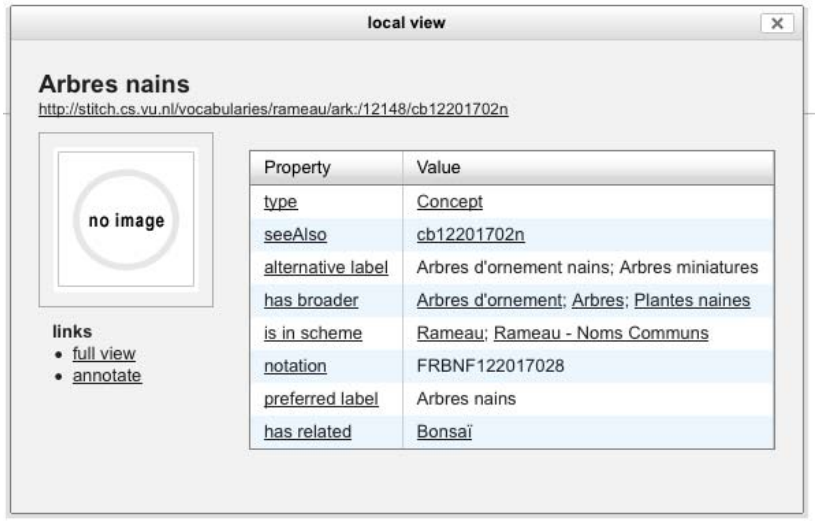


Fig. 3 Specific concept information

It is also possible to query the data using a SPARQL endpoint http://semanticweb.cs.vu.nl/europeana/sparql/.[4] 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 15 vocabularies. The following table gives a quantitative 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.[5]

[4] A more human-friendly SPARQL query input interface can be found at http://semanticweb.cs.vu.nl/europeana/user/query

[5] 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 211,854 instances of the skos:Concept class.

Table: Semantic Layer Contents

Vocabulary	Brief description
Cornetto[6]	Semantic network of Dutch word meanings, similar to Wordnet
Dutch AAT[7]	Dutch version of Getty's Art and Architecture Thesaurus
Getty AAT[8]	Art and Architecture Thesaurus
OSZK thesauri[9]	Thesauri at the National Library of Hungary
geothesaurus	Places
thesaurus	Subjects
RAMEAU[10]	Subject Thesauri at the French National Library
collectivités	Organisations
noms communs	Common nouns
noms géographiques	Places
personnes	Persons
subdivisions chronologiques	Time periods
titres	Titles of works
Systematik der Österreichischen Mediathek[11]	Thesaurus used at the Austrian media library
SCRAN classification[12]	Curriculum and topic classification of the SCRAN portal
Wordnet 2.0[13]	Princeton's semantic network of general English word meanings

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

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

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

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

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

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

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

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

Description of software developed for Europeana within EuropeanaConnect

Link to software	http://semanticweb.cs.vu.nl/europeana/session/thesaurus http://semanticweb.cs.vu.nl/europeana/browse/list_graphs
Login information	No username or password required
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	NA