ECP-2008-DILI-538025

JUDAICA Europeana

Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

Deliverable number: D2.5
Dissemination level: Public
Delivery date: 30 September 2011
Status: Final
Author(s): Dov Winer EAJC with EAJC, UB-FFM, AIU, MIBAC, Amitié, NTUA

This project is funded under the eContentplus programme\(^1\), a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

# Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>1.1 THE PURPOSE OF WORK PACKAGE 2</td>
<td>3</td>
</tr>
<tr>
<td>1.2 DESCRIPTION OF THIS DELIVERABLE</td>
<td>3</td>
</tr>
<tr>
<td>2. MOTIVATION</td>
<td>5</td>
</tr>
<tr>
<td>3. LIBRARIES</td>
<td>7</td>
</tr>
<tr>
<td>4. METADATA</td>
<td>10</td>
</tr>
<tr>
<td>5. LINKED DATA CONCEPTS</td>
<td>12</td>
</tr>
<tr>
<td>6. VOCABULARIES</td>
<td>16</td>
</tr>
<tr>
<td>7. SURVEY OF CONTROLLED VOCABULARIES IN THE THEMATIC DOMAIN</td>
<td>20</td>
</tr>
<tr>
<td>8. NAMES</td>
<td>21</td>
</tr>
<tr>
<td>9. PLACES</td>
<td>25</td>
</tr>
<tr>
<td>10. PERIODS AND TIME</td>
<td>29</td>
</tr>
<tr>
<td>11. HUBS OF JEWISH KNOWLEDGE</td>
<td>32</td>
</tr>
<tr>
<td>12. REPRESENTATION OF SELECTED CONTROLLED VOCABULARIES IN RDF/SKOS</td>
<td>34</td>
</tr>
<tr>
<td>12.1 RELEVANT VOCABULARIES AVAILABLE IN RDF/SKOS</td>
<td>34</td>
</tr>
<tr>
<td>12.2 REPRESENTING THE EJEWISH.INFO THESAURUS IN RDF/SKOS</td>
<td>34</td>
</tr>
<tr>
<td>12.3 REPRESENTING THE TAXONOMY - SYNOPTIC OUTLINE OF THE YIVO ENCYCLOPEDIA OF JEWS IN EASTERN EUROPE – IN RDF/SKOS</td>
<td>34</td>
</tr>
<tr>
<td>13. APPLICATION OF THE RDF/SKOS VOCABULARIES OF THE ISRAEL MUSEUM JERUSALEM FOR ACCESS TO EUROPEANA/JUDAICA EUROPEANA</td>
<td>36</td>
</tr>
<tr>
<td>14. REFERENCES</td>
<td>38</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 The purpose of Work Package 2

Judaica Europeana is selecting content related to the Jewish presence and heritage in the cities of Europe and will thus document the Jewish contribution to the European urban development. In cooperation with European cultural institutions Judaica Europeana will provide access to a large quantity of European Jewish cultural heritage at the level of the cultural object.

In this context, Work Package 2 of the Judaica Europeana project (WP2) is tasked with:

- Content identification and selection by means of auditing, assessing and selecting content to be digitised at the partner institutions collections and auditing in detail the available digitised resources. Establishing an advisory group of thematic domain experts that will support the process of content selection according to set criteria;
- Surveying the existing metadata schema used currently by the partners and facilitating the mapping of those standards to a common metadata standard;
- Assessing the requirements for the adoption of controlled vocabularies for Judaica purposes;
- Producing tools to support the conversion of the partners’ data into the common harvesting format for ingestion into the main Europeana service.
- Establishing a pilot knowledge management system to support the community of practice of scholars and cultural heritage professionals in the thematic domain area.

WP2 is in constant cooperation with other work packages in the project. In particular WP2 works closely together with WP3 and WP4: feeding information about standards for their work.

The present deliverable belongs to the following cluster of WP2 deliverables:

- D2.4 Survey of controlled vocabularies relevant to the thematic domain (M21)
- D2.5 Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS (M21)
- D2.7 Report on the deployment of the knowledge management system with a pilot focus group (M24)

The above listed deliverables jointly report on the completion of the following tasks:

T2.4 Controlled vocabularies survey, adaptation and semantic interoperability application

Identification and selection of existing controlled vocabularies in the thematic domain area, establishing and disseminating them throughout the domain:

- Adaptation of the selected vocabularies for JUDAICA purposes.

Adaptation of the chosen controlled vocabularies (taxonomies, thesauri and ontologies) for advanced indexing and retrieval of the content and to the semantic interoperability requirements defined for EUROPEANA (their representation in RDF/SKOS). With support of the technical WP3.

1.2 Description of this deliverable

An initial version of this document was prepared in September 2010, following the Europeana seminar in which the EDM – Europeana Data Model was presented at the Europeana V1.0 WG3.

Presented as an Internal Deliverable of Judaica Europeana that version played an important role in advancing the program it outlines. The document engaged the interest of many stakeholders and
enabled them to grasp better the vision that drives Europeana and seeks to establish a seamless universe of Cultural Heritage content embedded in the wider Linked Data web.

This paper outlines a program of actions concerning Jewish vocabularies. Its purpose is to enable the integration of content digitised by Judaica Europeana and many other related initiatives in the Web of structured data (Linked Data). This will enable access to Jewish knowledge much enriched by the professional communities that stored Jewish related expertise in databases and vocabularies.

This paper is structured in the following way:

We present a few examples derived from existing applications of Linked Data. Then we describe the radical changes that are happening concerning the role of the library catalogue. We follow this with some related observations concerning metadata. This will be followed with short sections introducing the main concepts related to the Linked Data: URIs, RDF, RDF Schema and SKOS for expressing vocabularies.

The core of this document is represented in the sections that outline a program of work concerning Jewish semantics. We explore current work concerning Names (Who?), Places (Where?) and Periods/Time (When); each such section concludes with a short outline of a possible Jewish-specific program of action in that area. The paper concludes with a review of additional vocabularies that may become hubs of knowledge in the Web of data when properly expressed in the required formats.

The goal of the Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources. The resulting structured Web can be queried through the SPARQL query language; crawled by RDF search engines, browsed by RDF enabled browsers. These tools feed innovative applications such as mashups that make use of such universal API.

The Linked Data approach emphasizes the re-use and linkage of richly described resources over the web. This is consonant to the Europeana Data Model ambition of making use of existing resources as well as supporting their enrichment, notably via the establishment of new relations between them. These resources may belong to one Europeana provider’s information space, to different providers’s spaces, or to external spaces used as knowledge references.

Readers that are already familiar with the Linked Data concepts may wish to go directly to the sections dealing with Jewish vocabularies concerning Names, Time, Place and Jewish hubs of knowledge.

---

1 Linked Data [http://linkeddata.org](http://linkeddata.org)
2 Linking Open Data [http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description](http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description)
3 Europeana Data Model Primer 05/08/2010 [http://tinyurl.com/edmprimer](http://tinyurl.com/edmprimer)
2. Motivation

The emerging structured Web – the Linked Data Semantic Web – will soon enable seamless access to an extraordinary extent of Jewish knowledge. A researcher working on matters related to the historian Heinrich Graetz will be able to find all his own publications and the works about him in different languages. The different ways his name is spelled (or misspelled) will not be an impediment due to the use of [controlled Names authorities]; Following the initial query he will be able to reach the manuscripts and other documents related to Graetz in the archive of the Jewish Theological Seminary of Breslau and others that are being digitised by the Jewish Historical Institute of Warsaw. Moreover, these will be seamlessly related to the documents in the Center for Jewish History in New York. Related gazetteers will enable the researcher to receive substantial information concerning every place that is referred in any standard biography of Graetz.

A tourist walking in Vilnius or Amsterdam will be able to point the camera of his cellular phone to a building or monument of interest and receive relevant aggregated information superimposed to the actual building. Such information may consist of pictures of the same building or monument in previous times or a multimedia presentation about it; a menu of references concerning the monument as they appear in digitised books or encyclopedias. To easily find people in the neighborhood that are ready to provide additional information. He will be able to add his own information, for example, his own pictures of the same monument that will be tagged and will immediately join other relevant information.

Entering a Jewish music digital repository will enable the user to get not only samples of the digitised song; links to places where it can be bought or downloaded and the lyrics - but also plenty of information about the composer, the genre, the period, the region in which it was composed, the different performances, reviews, historical background and more. These different pieces of information will come from disparate knowledge sources that have been published as Linked Data and integrated as a specific mash-up relating to the that music site. To see such kind of application visit the BBC Music Beta site1 built around Musicbrainz2 metadata and Dbpedia. Music metadata related artists are pulled from Musicbrainz and the introductory text for each artist's biography is taken from Dbpedia.

Jewish content holders will adopt what is becoming the dominant Web model, that of Linked Data and in time will publish their databases as RDF. However, in the Jewish realm, the realization of some of the possible scenarios described above depend on having in place support for authoritative replies to Jewish content related queries like: Who? Where? When? and support for well defined Jewish conceptual areas.

The wider technological developments do not guarantee this kind of support. It may be achieved only through concerted efforts by professional experts in the relevant content areas. The commitment of Jewish libraries, archives, museums, multimedia archives is thus essential. These efforts should be aimed at the identification of relevant vocabularies, their normalization and maintenance, their expression in the required formats, their publication in the new structured web and their application in new services that create seamless access to Jewish Knowledge.

The following paper seeks to review the basic concepts related to the new structured Web; to indicate some developments related to metadata; to outline the challenges that need to be tackled in

---

1 BBC Music Beta http://www.bbc.co.uk/music/
2 Musicbrainz http://musicbrainz.org/
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

areas like Names (who); Places (where); Time (when) and survey existing vocabularies that may together serves as hub in the Jewish Web of Knowledge.
3. Libraries

Libraries are well aware of the decline in the use of their catalogues. Students favour digital resources over the online library catalogue because such resources are available at anytime and from anywhere (Marcum, 2006). Calhoun (2006) in a report commissioned by the Library of Congress says "The existing local catalogues’ market position has eroded to the point where there is real concern for its ability to weather the competition for information seekers' attention" (p. 10). She proposes new uses for new users of the library catalogue and provides as examples mass digitisation, large scale integration with other systems, universal access.

Karen Coyle (Coyle, 2010) points out that the new library user no longer visits the physical library as his primary source of information. He seeks and creates information while connected to the global computer network. Libraries need to transform their public catalogue from a stand-alone database of bibliographic records to a highly hyperlinked data set that can interact with information resources on the World Wide Web. The library data can then be integrated into the virtual working spaces of the users served by the library. To become part of the dominant information system that is the Web the library catalogue should move from being “on the Web” to being “of the Web”. The linked data technology that has developed out of the semantic Web provides a path to follow. This is an ongoing process. Some implementations are available and there is ongoing work to adequate the main libraries standards, FRBR and RDA, to linked data best practices.

Implementation

Libraries have begun to publish critical sections of their catalogues as linked data. The German National Library (DNB)\(^1\) has published its person data (PND dataset describing 1.8 million people) and its subject headings (SWD; 164,000 headings) as Linked Data on the Web (see Hanneman and Kett, 2010). They have enriched their data by offering links to the German Wikipedia\(^2\) and Dbpedia\(^3\); to VIAF\(^4\); LCSH\(^5\) and RAMEAU\(^6\).

The Hungarian National Library published its entire OPAC and Digital Library as Linked Data\(^7\). The Library of Congress launched an experimental service that makes the Library of Congress Subject Headings available as linked-data using the SKOS vocabulary. LCSubjects.org takes the data made available from the Library of Congress' Authorities\(^8\) and is intended as a sandbox to provide relationships to other thesauri and web resources\(^9\).

A dramatic example of the new services that are becoming available is the Open Library project. Their goal is “One web page for every book ever published.” and already 20 million records are available. The basic design is of data elements composed by simple key/value pairs that can be re-

---

1 Linked Data Service of the German National Library  http://www.dnb.de/eng/hilfe/service/linked_data_service.htm
2 German Wikipedia  http://de.wikipedia.org
3 Dbpedia  http://wiki.dbpedia.org
4 VIAF – The Virtual International Authority File  http://viaf.org
5 Library of Congress Authorities  http://authorities.loc.gov
6 RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié)  http://rameau.bnf.fr
7 Hungarian National Library published its entire OPAC and Digital Library as Linked Data  http://lists.w3.org/Archives/Public/public-lod/2010Apr/0155.html
8 Library of Congress Subject Headings  http://id.loc.gov/authorities
9 Library of Congress Subject Headings as Linked Data  http://lcsubjects.org/
combined for a variety of uses. The individual units such as “author = John Smith” are available to be used as needed in whatever context is appropriate. Freed from a particular record structure the data is also available to link out to similar data in other data stores (Coyle, 2009). For example, any person named in the Open Library database can be linked to entries in Wikipedia for that person or to a personal web page.

Standards

**FRBR:** A general model of the library domain is provided by FRBR (Functional Requirements for Bibliographic Records). This is provided by the FRBR entities, relationships and attributes. The entities are presented in three groups. Group 1 represents the resource being described and has four entities: work, expression, manifestation and item. Group 2 represents agents that have relationships with Group 1 entities: persons, corporate bodies and families. Group 3 represents entities with a topical relationship to the Group 1 entities: concept, object, place and event and also Group 1 and 2 entities as they can be the subjects of any resource being described.

**RDA:** There is also a detailed set of data elements, vocabularies and guidance rules in library cataloguing standards, like the RDA (Resource Description and Access) whose final draft was published in 2009 - Cole (2010). The RDA has been defined in RDF using: the FRBR entities; the relations between entities as defined in RDA; and lists of terms that appear in the RDA document (value vocabularies following the Dublin Core Abstract Model Terminology). The description of RDA as RDF is available at the NSDL registry. The elements of the registry entry for properties are as follows: Identifier (URI) an identifier that begins with http://rdvocab.info/ identifying each term. Name – a machine friendly form of the name of the element. Label – a human-display label for the elements. Description – a human-readable definition as supplied in the RDA glossary. Domain – the class or classes to which the element belongs. The class is the FRBR entity with which the property is associated. Range - the value types that can be input as element contents. Type – the type of element, either property or sub-property, class or sub-class. subPropertyOf – for properties that have a hierarchical super ordinate property. hasSubProperty – for properties with sub-properties associated with them. The unique identifications of things and relationships assures that data can mix with other data without losing its specific meaning.

**FRBR ОО:** A related development is the publication of version 1.0.1 of FRBR ОО by the International Working Group on FRBR and CIDOC/CRM Harmonization. This is a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation and interchange of bibliographic and museum information. It applies empirical analysis and ontological structure to the entities and processes associated with works, to their properties, and to the relationships among them. Thereby it reveals a web of interrelationships which is also applicable to information objects in non-bibliographic arenas. (Bekiari, Doerr, Le Boeuf, 2010)

---

1 Open Library [http://openlibrary.org/](http://openlibrary.org/)
2 Dublin Core Abstract Model [http://dublincore.org/documents/abstract-model](http://dublincore.org/documents/abstract-model)
3 National Science Digital Library Metadata Registry [http://wiki.metadataregistry.org/Step-By-Step_Instruction](http://wiki.metadataregistry.org/Step-By-Step_Instruction)
4 The RDA (Resource Description and Access) Vocabularies [http://metadataregistry.org/rdabrowse.htm](http://metadataregistry.org/rdabrowse.htm)
The W3C established in 2010 the **Library Linked Data incubator group**\(^1\) with the purpose to increase global interoperability of library data on the Web. The group will explore metadata models, schemas, standards and protocols for building interoperability and library systems and networked environments.

\(^1\) W3C Library Linked Data Incubator group [http://www.w3.org/2005/incubator/lld/](http://www.w3.org/2005/incubator/lld/)
4. Metadata

Substantial changes occurred lately concerning ways to express and use metadata. Here we are concerned with the fact that metadata has become (1) machine actionable, (2) not contained in the limits of a record (3) able to become harmonized without becoming poorer (dumb down approach). Metadata is structured information that describes the attributes of information packages for the purposes of identification, discovery and sometimes management (Taylor, 2004). Metadata systems can be distinguished between two main categories. On the one hand structured metadata that uses a very basic template adequate for harvesting and interoperability among content repositories which themselves may have richer metadata structures. Such basic metadata enables resource identification and retrieval through federated repositories – popular examples of such kind of metadata are Dublin Core\(^1\), MuseumDAT\(^2\) and LIDO\(^3\) (developed lately for the museum community).

At another level of complexity we can find rich formats. Libraries, museums, archives, multimedia repositories communities have each developed comprehensive, detailed descriptions of their objects; these may combine metadata elements with encoding and controlled vocabularies. Examples are the MARC21\(^4\) schema for bibliographical records; the EAD\(^5\) for archives; CDWA\(^6\) and Spectrum\(^7\) for museums; MPEG-21 for multimedia\(^8\) ; LOM for learning objects\(^9\) etc. Such complex schemata in general include administrative, structural and descriptive metadata.

There is a disadvantage in the definition of common schemata of metadata for harvesting purposes. This process tries to harmonize richer metadata schemata and the result is information impoverishment. Systems that combine data from different sources using only the “dumb down” method reduces the metadata to the few matching elements and the result is the least rich metadata record possible (Coyle, 2009).

Rachel Heery and Manjula Patel (Heeri and Patel, 2000) introduced the idea of an 'application profile' as a type of metadata schema. They define application profiles as schemas which consist of data elements drawn from one or more namespaces, combined together by implementers, and optimised for a particular local application. Another approach to overcome the plurality of metadata schemata are crosswalks (Godby, Young and Childress, 2004). The online information environment produces a demand for compatibility with other descriptions: to locate materials in heterogeneous collections, to assemble a rich context for research or learning – and here crosswalks were intended to play a role. Despite the obvious differences two records belonging to different metadata schemata may have common elements. Both descriptions encode some understanding of author, title, and publisher. Though the correspondences are inexact, they are useful for promoting some degree of interoperability. We should note that metadata in both these approaches still remain confined in the structure of the record.

---

1 Dublin Core Metadata Initiative http://dublincore.org/
2 MuseumDAT - Harvesting Format for Providing Core Data from Museum Holdings http://www.museumdat.org
4 MARC21 Machine-Readable Cataloguing http://www.loc.gov/marc/
5 EAD –Encoded Archival Description http://libraries.mit.edu/guides/subjects/metadata/standards/ead.html
6 Categories for the description of works of Art http://www.getty.edu/research/conducting_research/standards/cdwa/
7 Spectrum terminology for museum staff http://www.collectionstrust.org.uk/spectrum-terminology/termgen
8 Moving Picture Experts Group (MPEG) http://mpeg.chiariglione.org/
Would it be possible to make use of all the relevant information available in the original metadata and still to achieve interoperability?

Harper (2010) remind us that historically, records – and not the statements about resources that they aggregate and package – have been treated as the central components of metadata. This was due to the attention paid how these packages are transmitted from one system to another. The problem with this conceptualization of metadata is that it arbitrarily limits the edges of descriptions to what can be effectively packages and transmitted in a record. Instead of focusing on the aggregation of individual pieces of metadata the Semantic Web community is advocating a focus on the smallest components of a resource’s description. This is now possible as the syntax of RDF made up of triples – statements composed of a subject, a predicate, and object where properties serve as predicates; the subjects are denoted by URIs defining the resources about which statements are made, and the objects – the value of the properties/predicate – can either be textual strings or additional resources. These statements can be linked together and woven into a rich tapestry of descriptions, forming a graph that extends its reach from myriad sources. In the “graph” paradigm it becomes easier to envision how library metadata interacts with other metadata on the open web. The unique identification of things and relationships assures that data can mix with other data without losing its specific meaning [see a more detailed description of these basic concepts below].
5. Linked Data Concepts

Linked Data is a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF\(^1\). Since 2007 it became a bandwagon – large databases are published every week - with more than 15 billion assertions now available.

Tim Berners-Lee summarized the main Linked Data guidelines as follows\(^2\):
- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provides useful information, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things.

The article “How to Publish Linked Data on the Web” (Bizer, Cyganiak and Heath, 2007) provide detailed instruction. Two recent articles provide a summative evaluation of the process, identify pending issues and providing reference to the software tools that can be employed: Bizer, Heath, Berners-Lee (2009); Hausenblas and Karnstedt (2010).

Here I will summarily describe the basic concepts involved:

**URI**

A Uniform Resource Identifier\(^3\) (URI) is a compact segment of characters that identifies an abstract or physical resource. The URI syntax defines a common grammar that is a superset of all valid URIs, allowing an implementation to parse the common components of a URI reference without knowing the scheme-specific requirements of every possible identifier. The term “resource” is used in a general sense for whatever might be identified by a URL – electronic resources as well as human beings, corporations and books; abstract concepts can be resources, as for example types of relationships (e.g., “parent” or “employee”) or numerical values. An identifier embodies the information required to distinguish what is being identified from all other things within its scope of identification.

While URIs may or may not be useful as locators in practice\(^4\), a URI scheme definition must be clear as to how it is expected to function. Schemes that are not intended to be used as locators should describe how the resource identified can be determined or accessed by software that obtains a URI of that scheme.

In many cases new URI schemes are defined as ways to translate between other namespaces or protocols and the general framework of URIs. For example the “ftp” URI scheme translates into the FTP protocol. For such schemes the description of the mapping must be complete, and in sufficient detail so that the mapping in both directions is clear: how to map from a URI into a set of protocol actions or name in the base namespace and how legal name values or protocol interactions might be represented in a valid URI.

---

1. Linked Data – Connect Distributed Data Across the Web [http://linkeddata.org/](http://linkeddata.org/)
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

As part of the definition of how a URI identifies a resource, a URI scheme definition should define the applicable set of operations that may be performed on a resource using the URI as its identifier.

HTTP URIs
The HTTP protocol is used by web servers and clients to request representations of Web documents and send back responses. HTTP includes capabilities to offer different formats and language versions of the same Web document and this process is known as “content negotiation”. This capability provides a recommended solution to the problem of access to identifiers (URI) of resources that are not documents such as persons, corporations, books, concepts. Such identifiers should be distinguished from representations of documents and photographs. This makes it possible assertions resident at distinct systems through the Semantic Web, that are made about the same object or concept, to be retrieved and processed, enabling the richness of services that may be offered.

The W3C Interest Group note of December 2008, “Cool URIs for the Semantic Web” recommends two approaches for making such references, Hash URIs and 303 Re-direction URIs. The following picture taken from that document nicely describes the procedure called Content Negotiation.

The URI for Alice the person is http://www.example.com/id/alice. The Web client that addresses this URI negotiates with the server; if the client is a human user using a regular browser the 303 redirect procedure leads him to a document that may be a description of Alice. If the client is a machine able to process RDF, it is redirected to an RDF representation of Alice that can, for example, fetched and integrated in the service that originated the lookout – a database or a mashup for example.

Namespace
Namespace is an abstract container providing context for the items (names, or technical terms, or words) it holds and allowing disambiguation of homonym items having the same name (residing in different namespaces). A namespace is also called a context, as the valid meaning of a name can change depending on what namespace applies. Names in it can represent objects as well as concepts. For many programming languages, a namespace is a context for identifiers. In an operating system, an example of namespace is a directory. It contains items which must have unique names.¹

¹ Namespace http://en.wikipedia.org/wiki/Namespace
RDF
RDF (Resource Description Framework) provides a way to express simple statements about resources, using named properties and values. It can be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web. RDF is intended for situations in which this information needs to be processed by applications, rather than being only displayed to people.

RDF is based on the idea of identifying things using Web identifiers (called Uniform Resource Identifiers, or URIs) and describing resources in terms of simple properties and property values. This enables RDF to represent simple statements about resources as a graph of nodes and arcs representing the resources, and their properties and values. This is represented in the form of subject-predicate-object expressions and are are known as triples in RDF terminology.

For example, one way to represent the notion "The sky has the color blue" in RDF is as the triple: a subject denoting "the sky", a predicate denoting "has the color", and an object denoting "blue". RDF is an abstract model with several serialization formats (i.e., file formats), and so the particular way in which a resource or triple is encoded varies from format to format.

A good tutorial on RDF was prepared by Ian Davis from Talis Research. The following example is taken from this tutorial.

In a relational database model the intersection of a row and a column in a table gives the value of a property (in the example below the Title) for a given thing (the book whose ISBN is 0596000480). The book has a title with a value of “JavaScript”

<table>
<thead>
<tr>
<th>isbn</th>
<th>title</th>
<th>author</th>
<th>publisherID</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>0596002637</td>
<td>Practical RDF</td>
<td>Shelley Powers</td>
<td>7642</td>
<td>350</td>
</tr>
<tr>
<td>0596000480</td>
<td>JavaScript</td>
<td>David Flanagan</td>
<td>3558</td>
<td>936</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

This is the essence of RDF: the (s,p,v) triple

---

1 RDF Primer – W3C Recommendation 10 February 2004 [http://www.w3.org/TR/rdf-primer/](http://www.w3.org/TR/rdf-primer/)


An (s,p,v) triple can be viewed as a labeled edge in a graph. Nodes in graph are things, arcs are relationships between things. Such graphs are amenable to graph algebraic operations like those described in the specifications of the RDF query language, SPARQL.

---

1 SPARQL Query Language for RDF – W3C Recommendation 15 January 2008 - [http://www.w3.org/TR/rdf-sparql-query/](http://www.w3.org/TR/rdf-sparql-query/)
6. Vocabularies

On the Semantic Web, vocabularies define the concepts and relationships used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. ¹ They support data integration by avoiding ambiguities that may exist between terms used in different data sets; vocabularies are used also to organize knowledge. A good example is the Book Vocabularies page at the W3C²

Knowledge Organization Systems (KOS) employ a variety of disparate terminologies in the form of term lists (e.g. authority files, glossaries, gazetteers, dictionaries), classification and categorization schemes (e.g. bibliographic classifications, taxonomies, categorization schemes) and relational vocabularies (e.g. thesauri, subject heading lists, semantic networks, ontologies). Terminology mappings (or vocabulary mapping) are essential to facilitate access and interoperability. It involves imposing equivalence, conceptual and hierarchical relationships between concepts in different schemes. The assumption underpinning mapping is that equivalence can exist between disparate knowledge organization systems and their respective terminologies - McCulloch and Macgregor (2008)

Defining RDF Vocabularies: RDF Schema³

As we have seen RDF provides a way to express simple statements about resources, using named properties and values. However user communities also need the ability to define the vocabularies (terms) they intend to use in those statements, specially to indicate that they are describing specific kinds or classes of resources, and will use specific properties in describing those resources.

The communities interested in describing bibliographic resources, for example, would want to describe classes such as ex2:Book or ex2:MagazineArticle and use properties such as ex2:author or ex2:title and ex2:subject to describe them.

RDF Schema provides the facilities to describe such classes and properties and indicate which classes and properties are expected to be used together. It allows resources to be defined as instances of one or more classes. It also allows classes to be organized in a hierarchical fashion; for example a class ex:Dog might be defined as a subclass of ex:Mammal which is a subclass of ex:Animal as well. The resources that belong to a class are called its instances.

User communities also need to be able to describe specific properties that characterize those classes of things (such as rearSeatLegRoom to describe a passenger vehicle). In RDF Schema, properties are described using the RDF class rdf:Property, and the RDF Schema properties rdfs:domain, rdfs:range, and rdfs:subPropertyOf.

RDF Schema also provides vocabulary for describing how properties and classes are intended to be used together in RDF data. The most important information of this kind is supplied by using the RDF Schema properties rdfs:range and rdfs:domain to further describe application-specific properties. The rdfs:range property is used to indicate that the values of a particular property are instances of a designated class. The instances of a range class are the possible objects (values) of a RDF (s,p,v) assertion.

¹ Vocabularies http://www.w3.org/standards/semanticweb/ontology
² Book Vocabularies http://esw.w3.org/BookVocabularies
³ RDF Primer W3C Recommendation 10 February 2004 http://www.w3.org/TR/rdf-primer
The rdfs:domain property is used to indicate that a particular property applies to a designated class. The instances of a domain class are used to indicate the subjects of a RDF (s,p,v) assertion. Many vocabularies that have been expressed as implementations of RDF Schema can be found in the Vocabulary Market\(^1\) and in the W3C page dedicated to Vocabularies and Ontologies.\(^2\) The W3C has also published Best Practice Recipes for Publishing RDF Vocabularies.\(^3\)

**SKOS**

SKOS—Simple Knowledge Organization System\(^4\)— provides a model for expressing the basic structure and content of concept schemes such as thesauri and other similar types of controlled vocabulary. It provides a simple way to express existing vocabularies in syntax adequate for their use in a Semantic Web, Linked Data context.

As an application of the Resource Description Framework (RDF), SKOS allows concepts to be composed and published on the World Wide Web, linked with data on the Web and integrated into other concept schemes.

**Concepts:** The fundamental element of the SKOS vocabulary is the concept that is identified with an URI; then it is asserted using the RDF property rdf:type that the resource identified by this URI is of type skos:concept using RDF triples:

```
  ex:animals rdf:type skos:Concept
```

**Labels:** SKOS provides three properties to attach labels to conceptual resources — skos:prefLabel, skos:altLabel and skos:hiddenLabel. They are all sub-properties of rdfs:label and are used to link a skos:Concept to an RDF plain literal which is a character string (e.g., “Love”) combined with an optional language tag (e.g., “en-US”).

Terms used as descriptors in indexing systems will be represented by skos:prefLabel. The skos:altLabel can be used for synonyms, near-synonyms, abbreviations and acronyms. Finally skos:hiddenLabel may be used to provide that a character string be available to applications performing text-based indexing and search operations but would not like that label to be visible otherwise. Hidden labels may be used to include misspelled variants of other lexical labels.

**Semantic Relationships:** The meaning of a concept is defined not just by the natural-language words in its labels but also, crucially, by its links to other concepts in the vocabulary. Mirroring the fundamental categories of standard compliant thesauri SKOS supplies three properties:

\(\text{skos:broader} \) and \(\text{skos:narrower} \) enable the representation of hierarchical links such as the relations between one genre and its more specific species, or the relationship between one whole and its parts. Example:

```
  ex:animals rdf:type skos:Concept;
  skos:prefLabel "animals"@en;
  skos:narrower ex:mammals.
```

\(^1\) Vocabulary Market \hspace{1em} \text{http://esw.w3.org/VocabularyMarket}

\(^2\) Common Vocabularies / Ontologies / Micromodels \hspace{1em} \text{http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies}

\(^3\) Best Practice Recipes for Publishing RDF Vocabularies, W3C Working Group Note, 28 August 2008 \hspace{1em} \text{http://www.w3.org/TR/swbp-vocab-pub/}

\(^4\) SKOS Home Page \hspace{1em} \text{http://www.w3.org/2004/02/skos/}
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

```
ex:mammals rdfs:type skos:Concept;
skos:prefLabel "mammals"@en;
skos:broader ex:animals.
```

`skos:related` enables the representation of associative (non-hierarchical) links. For example between an event and a category of entities which typically participate in it; or between categories where neither is more general or more specific.

```
ex:birds rdfs:type skos:Concept;
skos:prefLabel "birds"@en;
skos:related ex:ornithology.
```

```
ex:ornithology rdfs:type skos:Concept;
skos:prefLabel "ornithology"@en.
```

**Documentary Notes:** SKOS provides a `skos:note` property for general documentation purposes. Inspired by KOS standards such as ISO2788 and BS8723-2 this property is further specialized into `skos:scopeNote`, `skos:definition`, `skos:example` and `skos:historyNote`. In addition to these that are intended for users, SKOS includes two other specializations of `skos:note` useful for KOS managers or editors: `skos:editorialNote` and `skos:changeNote`.

**Concept Schemes:** in indexing practice, concepts usually come in carefully compiled vocabularies, such as thesauri or classification schemes. SKOS offers the means of representing such KOSs using the `skos:ConceptScheme` class.

The following example shows how to define a concept scheme resource (representing a thesaurus) and to describe that resource using the `dct:title` and `dct:creator` properties from Dublin Core [DC]:

```
ex:animalThesaurus rdfs:type skos:ConceptScheme;
dct:title "Simple animal thesaurus";
dct:creator ex:antoineIsaac.
```

Once the concept scheme resource has been created, it can be linked with the concepts it contains using the `skos:inScheme` property:

```
ex:mammals rdfs:type skos:Concept;
skos:inScheme ex:animalThesaurus.
```

In order to provide an efficient access to the entry points of broader/narrower concept hierarchies, SKOS defines a `skos:hasTopConcept` property. This property allows one to link a concept scheme to the (possibly many) most general concepts it contains:

```
ex:animalThesaurus rdfs:type skos:ConceptScheme;
skos:hasTopConcept ex:mammals;
skos:hasTopConcept ex:fish.
```

**Networking Knowledge Organization Systems (KOS) – Mapping Concept Schemes:**

When concepts from different concept schemes are connected together they begin to form a distributed, heterogeneous global concept scheme. Such web of concept schemes can allow meaningful navigation between KOS.

Every SKOS concept is assigned a URI. This is useful for establishing semantic relations between pre-existing concepts. Such mappings are crucial for applications that use several KOS at the same time, where these KOS have overlapping scopes and need to be semantically reconciled.

18/39
A crucial feature of mapping is the possibility to state that two concepts from different schemes have comparable meanings, and to specify how these meanings compare, even though they come from different contexts and possibly follow different modeling principles. Conceptual mappings are the key advantage of making KOS available on the Semantic Web using SKOS.

SKOS provides several properties that map concepts between different concept schemes. This can be done by asserting that two concepts have a similar meaning, using the `skos:exactMatch` and `skos:closeMatch` properties.

Two concepts from different concept schemes can also be mapped using properties that parallel the semantic relations introduced previously: `skos:broadMatch`, `skos:narrowMatch` and `skos:relatedMatch`.

It is possible to map the concepts in `ex1:referenceAnimalScheme` to the concepts in `ex2:eggSellerScheme` by using the mapping assertions below:

```
ex1:platypus skos:broadMatch ex2:eggLayingAnimals.
ex1:platypus skos:relatedMatch ex2:eggs.
ex1:animal skos:exactMatch ex2:animals.
```

One example of the possibilities open by the possibility of mapping of concepts across schemes is to establish connections between well maintained and established terminologies with those developed locally by small and medium institutions. Such mapping enable the establishment of a catching net for information base in well established knowledge bases.

---

1 SKOS Implementation Report May 19th 2009

[http://www.w3.org/2006/07/SWD/SKOS/reference/20090315/implementation.html](http://www.w3.org/2006/07/SWD/SKOS/reference/20090315/implementation.html)
7. Survey of Controlled Vocabularies in the Thematic Domain

The Judaica Europeana deliverable D2.4 Survey of Controlled Vocabularies in the Thematic Domain is being provided at the same time as the present one. It provides descriptions of dozens of such vocabularies grouped according to their main focus: Names, Places, Periods, Vocabularies of Broad interest, Vocabularies of local interest. In the next paragraphs of the present document we show how some of these vocabularies may be applied in an EDM/Linked Data context and suggest a program of actions to be taken in concert by the relevant stakeholders concerning Names, Places, Periods and establishing Hubs of Knowledge through vocabularies of broad interest in the thematic domain. The suggested work in the area of Jewish content is preceded by a discussion of vocabularies of general interest in these areas.

Additional multilingual vocabularies were described a few years ago in the survey carried out in Israel by a working group coordinated by Dr. Allison Kupietzky from the Israel Museum Jerusalem as part of the MINERVA project. See: [http://mek.oszk.hu/minerva/html/dok/israel.doc](http://mek.oszk.hu/minerva/html/dok/israel.doc) For other countries see: [http://mek.oszk.hu/minerva/html/publikaciok.htm](http://mek.oszk.hu/minerva/html/publikaciok.htm)
8. Names

Names – of persons, authors, musicians, painters, families, corporations, institutions – are one of the most appropriate for indexing and retrieving information. Many communities are involved in collecting, storing, managing and maintaining directories of names. Governments, phone companies, Internet providers, publishers, rights clearance institutions, archives, national libraries and many more carry out these tasks. A good collection of resources concerning this matter can be found in the NISO webinar on name identifiers¹

Robert Wolven says that as web discovery transcends the boundaries surrounding communities and aggregations of information, the range of names encountered becomes ever larger and more diverse, and a broader framework for identification is of vital importance. New problems demand new solutions, and may change the ways we think about identity and authority.²

Some efforts related to our concerns:

FOAF - FOAF (an acronym of Friend of a friend) is an ontology describing persons, their activities and their relations to other people and objects. Anyone can use FOAF to describe him or herself. FOAF allows groups of people to describe social networks without the need for a centralised database. FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF). Computers may use these FOAF profiles to find, for example, all people living in Europe, or to list all people both you and a friend of yours know. Each profile has a unique identifier (such as the person's e-mail addresses or a URI of the homepage or weblog of the person), which is used when defining these relationships.³ One can register through FOAF-a-Matic⁴ service; and a FOAF namespace document⁵ is available from January 2010.

MusicBrainz captures information about artists, their recorded works, and the relationships between them. Recorded works entries capture at a minimum the album title, track titles, and the length of each track. These entries are maintained by volunteer editors who follow community written style guidelines. As of 5 April 2010 (2010-04-05)[update], MusicBrainz contained information about 533,145 artists.⁶

ISAAR (CPF) is a companion standard of ISAD(F) General International Standard Archival Description. It provides guidance for preparing archival authority records which provide

1 NISO Name Identifiers Webinar Resources http://www.niso.org/news/events/2010/nameid/resources
4 FOAF-a-Matic http://www.ldodds.com/foaf/foaf-a-matic
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

descriptions of entities – corporate bodies, persons and families associated with the creation and maintenance of archives.¹

Names authorities in libraries: traditionally the main function of the library catalogue is to enable to find which works by a given author are in the library. However, lately, the Statement of international cataloguing principles² issued by IFLA in February 2009 recognizes that the computer and the network are not bound by the constraints of the card catalogue and enounced the principle of comprehensiveness. It requires including any person, family or corporate body associated with a given resource.

Alan Danskin ³ reviews critically the implications of the principle of comprehensiveness. He indicates that present arrangements for the maintenance of authority control of names in National Libraries are inadequate. They are restricted to names associated with books while the preferred means by which academics and researchers publish their work are articles. Even in the humanities increasingly preprints, conference proceedings and web resources are the currency of research. The scales on which journal articles and web pages, let alone digitised documents are produced far exceed the capabilities of current workflows. He suggests that authority control moves from being a craft to an industry responsibility. There is a lot of publisher interest in accurate identification of authors, particularly to track rights; grant making bodies and higher education institutions are also increasingly interested in unambiguous identification of who has written what in order to measure performance. He adds that there is an opportunity and incentive to put the onus on providing/recording adequate personal metadata to the author/publisher.

VIAF (The Virtual International Authority File)
VIAF is a joint project of several national libraries, implemented and hosted by OCLC. The project's goal is to lower the cost and increase the utility of library authority files by matching and linking the authority files of national libraries, and then making that information available on the Web. Users are able to see names displayed in the most appropriate language. Users in their respective countries will be able to view name records as established by the other nations, thus making the authorities truly international and facilitating research across languages anywhere in the world.⁴ For instance, it should be possible to search for Twain, Твен or טוין, retrieve materials about Mark Twain and be able to read the names associated with them in your preferred script.

Thomas Hickey⁵ explains that while VIAF's target audience is librarians who deal with international materials, our goals for enhanced searching overlap with those of the Semantic Web. To support this, VIAF is available as linked data, supporting machine as well as Web browser access. He explains that they make the links that form the basis of the virtual authority file by collecting personal name authority records and their associated bibliographic metadata. This lets us match

names not only on the name itself and any cross-references in the authority records, but to also use information found in bibliographic records about which works a person has written. The enriched VIAF records created as the result of all this matching bring together more information than exists in any single authority record.

In early 2010 VIAF has over 10 million personal names in it derived from nearly 13 million authority records from 20 different files. To support the matching we are also managing some 70 million bibliographic records which we match against the authorities and extract additional information (e.g. titles, coauthors, publishers) that can be associated with names.

VIAF is already available as Linked Data Their new linked data announce in May 2010 retain the SKOS description but also describe the VIAF concept as FOAF and expose more of the VIAF data in a more 'native VIAF' form. We do this by minting some new URIs that the RDF describes: Name Authority Cluster; as SKOS Concept; as FOAF Person. They also provides several views:

**ISNI - International Standard Name Identifier**

ISNI is a new identifier standard for names. It will provide a means to uniquely identify the publicly facing names of authors, composers and other creators, fictional and historical characters and rights holders, particularly publishers. Such an authoritative identifier will serve as a link for occurrences of an identity across databases on the Web and make it easier to relate names used by publishers to those used in libraries.

The ISNI is expected to become operational in 2010. ISNIs will be initially assigned by matching records supplied from the consortium members. They have successfully conducted a series of tests using the VIAF file and its underlying matching processes. It is anticipated that around 3 million ISNIs could be assigned and ready for diffusion from day 1. See its FAQ and site.³

**Access to relevant Jewish related information – implications from the review in the area of Names**

**Short term tasks**

1. The VIAF authorities include the Library of Congress and the National Library of Israel. These files guarantees a substantial representation of names of authors and other relevant persons related to Jewish content.
2. VIAF is already available as Linked Data
3. The immediate task is to disseminate the use of these VIAF URIs. Any institution maintaining Jewish related information resources that substitutes the name strings by the VIAF URIs will ipso facto extraordinarily extend the reach of information that can be accessed. Minimally this includes the different spellings for the same name; links provided by VIAF to the works related to these names; information about publishers and more.

**Medium term tasks:**

1. Alan Danskin (see above) indicates the names authority files produced by national libraries do not include periodicals and so are incomplete leaving outside its scope a critical area.

---


² Thomas Hickley. VIAF’s new linked data (May 27, 2010) [http://outgoing.typepad.com/outgoing/2010/05/viafs-new-linked-data.html](http://outgoing.typepad.com/outgoing/2010/05/viafs-new-linked-data.html)

2. We should strive to include names derived from periodical publications related to Jewish content to be included in the Names Authority File. RAMBI, the main bibliographic initiative concerning Judaica should become involved in this effort.

3. RAMBI - The Index of Articles on Jewish Studies - is a selective bibliography of articles in the various fields of Jewish studies and in the study of Eretz Israel. Material listed in Rambi is compiled from thousands of periodicals and from collections of articles - in Hebrew, Yiddish, and European languages. The main criterion for inclusion in the bibliography is that the article be based on scientific research, or contain important information for such research. The editorial board has striven to include in it all of the important articles published throughout the world in the field of Judaica.

Long term tasks:
1. Current work flows in the management of authorities are unable to accomplish the IFLA principle of comprehensiveness. There is need to involve publishers, grant making bodies, higher education institutions, rights clearance associations – constituencies with a vested interest in the accurate identification of authors. One example of such kind of initiative is the JISC Names project¹ which seeks to uniquely identify individuals and institutions involved in research in higher education in the United Kingdom.

2. Estimation of the feasibility of a concerted effort targeting the communities with an interest in Jewish content should be assessed. Or alternatively, to follow larger developments in this area and prepare the tools for their early adoption when the process matures.

¹ Names project (JISC)  http://names.mimas.ac.uk/
9. Places

Locations related to Jewish life need to be identified in their variety of names, spellings, languages and historical context. A summary of initiatives in this area follows below. We conclude with an outline of next steps to be taken concerning Jewish places names.

Becker and Bizer (2009) define the Geospatial Semantic Web that results from applying Linked Data principles to geographical information. It enables to interlink data about locations between data sources and to relate locations to each other, as well as to further Web content using explicit typed links, in addition to geographic coordinates. For instance, a location could be linked to its encompassing locations in an administrative hierarchy, as well as to persons who were born, died or worked there. Semantic Web clients may then navigate across these explicit links to retrieve data describing the interlinked entities.

Here we list some data sources for geographical information:

**NGA GEOnet Names Server (GNS)** The Geographic Names Server is the official repository of standard spellings of all foreign place names, sanctioned by the United States Board on Geographic Names. The database also contains variant spellings (cross-references), and is starting to hold the native script spellings of these names. All the geographic features in the database contain information about location, administrative division, and quality. The coordinate system for data served by GNS is WGS84. Coordinates in the GEOnet Names Server are approximate and are intended for finding purposes only.

**Open Geospatial Consortium (OGC)** The OGC is an international, voluntary consensus standards industry organization that is leading the development of standards for geospatial and location based services. It includes 399 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC standards can be used to geospatially enable interoperable Web based applications and portals. These applications or portals can provide either free or available-for-fee services and content that is widely available to Web users.

Revisions to the OGC naming policy allow for OGC names structured as http URIs, as an alternative to URNs. The use of http URIs (a) resolves some deployment challenges and (b) provides an opportunity for easier engagement with broader communities. So OGC should now consider taking the next step, and mandate the use of http URIs for persistent identifiers in OGC standards. The OGC Naming Authority Policy Documents define the rules for defining OGC names.

**GeoNames** The GeoNames geographical database provides data such as names in different languages, feature type and geo-coordinates for over 8 million places. The predicate `geoNames:parentFeature` is used to link to a resource’s parent within GeoNames, resulting in a feature hierarchy that maps administrative subdivisions of a country and links countries to a

---

5. GeoNames [http://www.geonames.org/about.html](http://www.geonames.org/about.html)
continent which are linked to a single Earth resources. The Linked Data interface further provides pointers to retrieve a resource’s children feature, nearby features, and for countries, neighboring countries. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes¹. GeoNames is 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). Users may manually edit, correct and add new names using a user friendly wiki interface. The data is accessible free of charge through a number of webservices and a daily database export. GeoNames is already serving up to over 11 million web service requests per day.

**Freebase**² Freebase is an online database which users can edit in a similar fashion as they edit Wikipedia articles. It contains data about 1,751,020 locations which is served as Linked Data on the Web. The location commons holds position information for topics, as well as political entities like countries, administrative divisions (states, provinces, departments, etc.) and cities. The political portion of this domain is currently undergoing a revision, with the ultimate goal of distinct types for the major administrative divisions of each country.

**LinkedGeoData**³ An effort to add a spatial dimension to the Web of Data / Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative. Its data structure is composed of individual nodes that represent individual points, and ways that form collections of nodes, such as roads and rivers. The Linked Data output indicates nearby features. Online access⁴ is provided via a SPARQL-Endpoint and a LinkedData-Interface. The community project LinkedGeoData was started and is administered by the AKSW research group from the University of Leipzig. One important move (July 2010) has been their decision to pen source the code for the LinkedGeoData project and move it to Google Code: [http://code.google.com/p/linkedgeodata/](http://code.google.com/p/linkedgeodata/).

**Geo URI**⁵ The IETF (Internet Engineering Task Force) published on June 2010 the RFC 5870 A Uniform Resource Identifier for Geographic Locations (‘geo’ URI). This document specifies a Uniform Resource Identifier (URI) for geographic locations using the ‘geo’ scheme name. A ‘geo’ URI identifies a physical location in a two- or three-dimensional coordinate reference system in a compact, simple, human-readable, and protocol-independent way. The default coordinate reference system used is the World Geodetic System 1984 (WGS-84).

**TGN (The Getty Thesaurus of Geographic Names)®**⁶ The TGN contains around 895,000 records, including around 1,115,000 names, place types, coordinates, and descriptive notes. It is currently published in both a searchable online Web interface and in data files available for licensing. It is attended potential users of geographic vocabulary in cataloging and scholarship of art and architectural history and archaeology.

---

¹ GeoNames features [http://www.geonames.org/export/codes.html](http://www.geonames.org/export/codes.html)
³ LinkedGeoData [http://linkedgeodata.org/About](http://linkedgeodata.org/About)
⁴ LinkedGeoData [http://linkedgeodata.org/About](http://linkedgeodata.org/About) Online access: [http://linkedgeodata.org/OnlineAccess](http://linkedgeodata.org/OnlineAccess)
⁶ TGN (Thesaurus of Geographic Names) [http://www.getty.edu/research/conducting_research/vocabularies/tgn/about.html](http://www.getty.edu/research/conducting_research/vocabularies/tgn/about.html)
Its scope includes terminology needed to catalogue and retrieve information about the visual arts and architecture; it is constructed using national and international standards for thesaurus construction; it comprises a hierarchy with tree structures corresponding to the current and historical worlds; it is based on terminology that is current, warranted for use by authoritative literary sources, and validated by use in the scholarly art and architectural history community; and it is compiled and edited in response to the needs of the user community. The minimum record for each place includes a name, a place type, and a position in the hierarchy that shows its parent places (or broader contexts). The name alone does not identify a place because there may be many homographs. While many records in TGN include coordinates, these coordinates are approximate and are intended for reference ("finding purposes") only.

Dates for the Names: Dates comprise a Display Date, which is a note referring to a date or other information about the name, and Start Date and End Date, which are years that delimit the span of time referred to in the Display Date. Start and End Dates index the Display Date for retrieval, but are hidden from end-users (for Siena, Italy, start date based on life dates of Julius Caesar) Saena Julia (H,V,N,Latin-P)…. Roman (start:1\-100, end: 300)

Relationships in the TGN: The TGN includes equivalence, associative, and hierarchical relationships. (1) **Equivalence Relationship.** All relationships between names within the same TGN record are *equivalence relationships*. Among all the names that refer to the place, one is indicated as the *preferred name,* . This is the *vernacular* or local-language name most often found in scholarly or authoritative published sources. The preferred English name is also indicated. Variant and alternate names in the record include names in other languages, names transliterated into the Roman alphabet by various methods, names in natural or inverted form (particularly for physical features, e.g., *Etna, Mount*), nicknames, official names, and historical names. Misspellings may be included if they are found in published sources. (2) **Hierarchical Relationship.** The *hierarchy* in the TGN refers to the method of structuring and displaying the places within their broader contexts. *Hierarchical relationships* in TGN represent part/whole relationships. TGN is *polyhierarchical*, meaning that a place may have multiple parents or broader contexts. For example, the US state of Hawaii is administratively part of the United States in North America, but it is physically located in Oceania. (3) **Associative Relationship.** *Associative relationships* may exist between the records for places in TGN. For example, if an inhabited place has been physically moved (as when the location has been deemed unsafe due to flood or earthquake), there should be an associative relationship between the original settlement and the new settlement.

**Jewish related information – implications concerning Places**

The FamilySearch Wiki provides a good review of gazetteers relevant for Jewish genealogy. When learning about a locality for genealogical purposes, you should use both old and modern gazetteers. Old gazetteers have information about older jurisdictions, Jewish communities that no longer exist, and town names that have changed over the years. Some names have changed several times as the boundaries and governments of a country have changed, and the name may be different in family documents from how it is listed today. On the other hand, modern gazetteers are also important for genealogical work. They can be used to determine how the town name is spelled today, which may be crucial for finding the town on a map. It is necessary to know how the town name is spelled today and where it is located in order to write letters requesting records.
**JewishGen** includes among its services a gazetteer\(^1\) with 350,000 towns in 24 countries in Central and Eastern Europe. It is based on the Geographic Names Database (GNDB) compiled by the U.S. Defense Mapping Agency, which was also used extensively in the compilation of Where Once We Walked. It has links to maps showing where various towns are located in Europe. This system searches by the Daitch-Mokotoff Soundex which help find a town name even if it is spelled slightly differently from the gazetteer.

**Yad Vashem** developed and maintains The Central Database of Shoah Victims' Names\(^2\). The database includes a table for Locations that in fact is one an important Gazetteer for Jewish related place names.

**Short term tasks related to Jewish places information for Linked Data:**

1. Encourage JewishGen and Yad Vashem to publish their gazetteers as linked data.

**Medium term tasks related to Jewish places information for Linked Data:**

1. Establish a systematic registry of Jewish places relevant gazetteers.

2. Check the feasibility of expressing them in RDF.

3. Check the feasibility of establishing a community based Jewish places gazetteer service. It could be based on successful prototypes like those developed by the GeoNames, the Freebase or the LinkedGeoData projects.

---

\(^1\) JewishGen Gazetteer [http://www.jewishgen.org/ShtetlSeeker/loctown.htm](http://www.jewishgen.org/ShtetlSeeker/loctown.htm)

\(^2\) Central Database of Shoah Victim’s Names [http://www.yadvashem.org/wps/portal/IY_HON_Welcome](http://www.yadvashem.org/wps/portal/IY_HON_Welcome)
10. Periods and Time

The categorization of time into discrete named blocks is called periodisation. It plays an essential role in the documentation and data retrieval in arts, music, history, archeology and more. Some tools and approaches for the establishment of consistent periods are reviewed here. We conclude with the indication of some tasks concerning periodisation in the context of Jewish concerns.

Doerr, Kritsotaki and Stead (2008) make a distinction between simple and complex approaches for chronological reasoning. There are statistical approaches to date archeological strata that regard the studied phenomena as simple, well-defined and associated with precise time points. On the other hand the notion of cultural periods in archeology is based on cultural semantics. Chronological systems based on such periods are notoriously controversial, due to the complexity of the relationships between contextual phenomena and spatiotemporal values.

CIDOC CRM
It defines the basic notion of a period as: “this class comprises sets of coherent phenomena or cultural manifestations bounded in time and space. It is the social or physical coherence of these phenomena that identify an E4 Period and not the associated spatio-temporal bounds. These bounds are a mere approximation of the actual process of growth, spread and retreat. Consequently, different periods can overlap and coexist in time and space, such as when a nomadic culture exists in the same area as a sedentary culture… “. (Crofts, Doerr, Gill, Stead and Stiff 2004)

Getty Art and Architecture Thesaurus
The Art & Architecture Thesaurus (AAT)\(^1\) is a controlled vocabulary used for describing items of art, architecture, and material culture. Thousands of AAT terms are added and edited every year. Currently the AAT contains around 34,000 records for concepts, including 131,000 terms, plus descriptions, bibliographic citations, and other information.

The AAT is a faceted classification system as well as a hierarchical one. There are seven facets:

- Associated Concepts - abstract concepts;
- Physical Attributes - perceptible or measurable characteristics such as size, shape, chemical properties, texture and hardness;
- Styles and Periods - stylistic groupings and distinct chronological periods;
- Agents - people, groups of people, and organizations;
- Activities - areas of endeavor, physical and mental actions or methods;
- Materials - physical substances;
- Objects - objects either fabricated or given form by human activity.

The AAT can be used in three ways: at the data entry stage, by catalogers or indexers who are describing works of art, architecture, material culture, archival materials, visual surrogates, or bibliographic materials; as knowledge bases, providing information for researchers; and as search assistants to enhance end-user access to online resources.

Thesauri of historical periods
Martin Doerr and his colleagues (Doerr, Kritsotaki and Stead (2008) developed a methodology to create thesauri of historical periods. The main distinction they make is between phenomena that have left distinct traces and are taken as objective indicators for the coherence of the respective period (the identity criterion), and other characteristic phenomena, distinct or not, that are either product of interpretation or that are not directly associated with the coherence of the period as a whole.

They bring as an example “Ming Dynasty” that is defined by the political system – any change of

\(^{1}\) The Getty Art & Architecture Thesaurus (AAT)
http://www.getty.edu/research/conducting_research/vocabularies/aat/faq.html
our knowledge about the dates of rule of the Ming emperors will change the asserted temporal bounds of “Ming dynasty”. On the other hand Ming porcelain is a good indicator for dating finds from the Ming dynasty. However change in our knowledge of the dates of production of a certain Ming porcelain style will not affect the temporal bounds of Ming, but at most characterize this style as no more central to Ming etc. In this case the political system provides the identity criterion, and the porcelain production just a distinct, characteristic phenomenon of this period.

In contrast to the identity criteria, general characteristics of a period can be documented as part of their description, such as technological activities, social-political structures, economy and trading, history of war activities, patterns of settlements and belief systems, generally different aspects of material culture. They do not define, but simply describe and interpret a period.

As part of the Schema: Terminology and Definition they distinguish between: Time period [Technology use; Style based; Stratum based; Dynasty or ruling period; Sociopolitical system; Cultural influence]; Starting event, Type of event [ Technological inventions; Imports or cultural borrowings; Natural catastrophes; Abandonment; Destruction; Religious event; Social political event]

As part of the Schema: Spatiotemporal Extent they distinguish between Spatiotemporal relationships: Spatiotemporal [falls within; containing; overlaps with; separated from]; Temporal [finishes; is finished by; starts; is started by; includes; occurs during; overlaps in time with; is overlapping in time by; meets in time with; is met in time by; occurs before; occurs after; is equal in time to]; Spatial [ Falls within place of; consists of place of; forms part of place of; contains place of; overlaps place of; borders with place of].

Jewish related information – implications concerning Periods

There are several initiatives in Jewish related contexts to cope with issues of periodisation of Jewish history and cultural production. Such works are deeply embedded in the catalogues and indexing systems of institutions like the Israel Museum Jerusalem1; Israel Antiquities Authority2; the YIVO Institute for Jewish Research3; the Center for Jewish History4 and many more.

A directory of historical timelines is maintained by the Jewish History Resource Center5 of the Dinur Center for Research in Jewish History at the Hebrew University of Jerusalem. Some of such timelines should be further studied to consider their potential as periods indexing tools to be converted.

An excellent example for guidelines on scholarly classification that includes periodisation elements is the “Comparative Development of the Classes for Religious Law: The Abrahamic Tradition”6 prepared by Jolanda Goldberg from the Library of Congress.

The Jewish Museum Berlin developed a thesaurus for German Jewish History. The project was carried out by Dr. Iris Blochel-Dittrich. This thesaurus is managed in the framework of BAM the German Libraries, Museums and Archives consortium and is already expressed in SKOS using the

1 Israel Museum Jerusalem http://www.imj.org.il
2 Israel Antiquities Authority http://www.antiquities.org.il/home_eng.asp
3 YIVO Institute for Jewish Research http://www.yivo.org/
4 The Center for Jewish History http://www.cjh.org/
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

xTree software package that supports collaborative work in the maintenance of controlled vocabularies.

**Short term tasks:**
1. Survey available vocabularies for periodisation of Jewish history and culture
2. Seek to express some of such vocabularies as RDF Schema/SKOS and publish them as Linked Data
3. Promote the interlinking of such vocabularies with Jewish content resources

**Medium and long term:**
1. Establish the institutional tools enabling an in-depth probing of present practices for Jewish history and culture periodisation.
2. Review available approaches like those briefly described above.
3. Consider the development of pilots experiments implementing such approaches in Jewish history and culture.
11. Hubs of Jewish Knowledge

Some outstanding initiatives have developed extensive indexing systems that can be considered as hubs of Jewish knowledge. They produced corpora of well structured knowledge bases whose purpose is the description of Jewish related objects.

The publication of such databases as Linked Data in the RDF Semantic Web will lead to an increased application of such knowledge assets enriching their value both for the users as for the institutions that developed them. Here we bring short descriptions of some of such potential hubs.

**The YIVO Encyclopedia of Jews in Eastern Europe¹**

The YIVO Encyclopedia of Jews in Eastern Europe has been planned to be the definitive reference work on all aspects of the history and culture of Jews in Eastern Europe from the beginnings of their settlement in the region to the present. This two-volume, 2,400-page encyclopedia, comprising approximately 2 million words, almost 1,200 images (including two 16-page color inserts), and 55 maps, draws on the most current scholarship in all relevant fields and explores Jewish life in all its variety and complexity. The editors’ goal has been to cover everything of cultural or historical significance using an ecumenical, nondenominational, and non ideological editorial approach. **This project is unprecedented. To this day, a full-fledged encyclopedia dedicated exclusively to the centuries-long history and culture of East European Jewry has never appeared.**

Under the direction of Editor in Chief Gershon David Hundert, professor of history and chair of the Department of Jewish Studies at McGill University, 450 internationally recognized scholars has served as editors and contributors. *The YIVO Encyclopedia* not only provides a forum for their collective knowledge, but also serves as a meeting point for a new generation of scholars from former Communist Europe and their colleagues from North America, Israel, and Western Europe. The encyclopedia brings their scholarship together for the first time.

This project have two crucial features for our present context.

1. The encyclopedia was published online and all its contents are available at [http://www.yivoencyclopedia.org/](http://www.yivoencyclopedia.org/)

Expressing these resources in RDF and publishing it as Linked Data will establish them as the core of the knowledge available as Linked Data. The encyclopedia entries will become much enriched with primary sources. The documents of the community of Breslau, including those of the Seminary in which Graetz taught will for example be directly available from the relevant entries in the Encyclopedia and reversely will lead to the relevant entries for further explanations about their context of creation and use.

**Sfardata²**

"Sfardata: The Codicological Database of the Hebrew Palaeography Project, Israel Academy of Sciences and Humanities, Jerusalem – A Tool for Historical Typology, Dating and Localizing Medieval Manuscripts" by Malachi Beit-Arié.

SfarData is a sophisticated quantitative database and retrieval system of a large number of

---


measurable codicological attributes recorded in most of the extant explicitly dated, and in the undated but otherwise 'colophoned,' or named, Hebrew medieval manuscripts. Sponsored by the Israel Academy of Sciences and Humanities, in collaboration with the Jewish National and University Library in Jerusalem, in cooperation with the Institut de Recherche et d'Histoire des Textes in Paris, The Hebrew Palaeography Project has been engaged since its inception in 1965 in studying and recording most of the visible and quantifiable codicological features and variables of all the surviving dated Hebrew codices and the undated ones with indications of scribes' names, some 5500 manuscripts kept in collections all over the world. Since the early 1970's many of the attributes recorded in the detailed questionnaires have been coded and electronically stored in Jerusalem. Over the past ten years an elaborate retrieving, sorting and linking system, based on FoxPro software (recently converted into Window-base system, powered with Visual FoxPro 5, including also images of the manuscripts), was and is still being developed, allowing endless querying of the data, clustering and statistics1.

Center of Jewish Art at the Hebrew University of Jerusalem - The Jerusalem Index of Jewish Art
For nearly thirty years the Center for Jewish Art, a research institute at the Hebrew University of Jerusalem, has carried out a comprehensive documentation program of the visual cultural heritage of the Jewish people. Through detailed descriptions and photography by our researchers the vanishing objects are preserved for future generations. Prof. Bezalel Narkiss, Israel Prize laureate, established the Jerusalem Index of Jewish Art, in which there are at present about 200,000 documents of objects originating from all over the globe, which range from a coin to a complex of synagogues. Documentation is carried out in five sections of the Index: Ancient and Modern Arts, Hebrew Illuminated Manuscripts, Synagogues and Ritual Objects, as well as Jewish Ritual Architecture. The objects are also classified according to iconographical subjects, with references to textual sources for cross-reference, biographies and bibliography. Examples of documentation:
Ritual Objects http://cja.huji.ac.il/Seker.html
Illuminated Manuscripts http://cja.huji.ac.il/Hebrew_Illuminated_Manuscripts.html
Ancient Jewish Art http://cja.huji.ac.il/atika.html
Modern Jewish Art http://cja.huji.ac.il/modern.html
Ritual Architecture http://cja.huji.ac.il/architecture.html

IMAGINE – The Israel Museum Collections Database Project2
The IMAGINE Thesaurus was developed and is used by the Israel Museum, Jerusalem, an encyclopaedic museum, with standards garnered from the VRA and the AAT, focused mainly on Jewish material culture. It is constructed of "legacy terms" and is multidisciplinary in its nature. The Israel Museum has benefited from the Israel Antiquities Authority lexicon, and has continued to work on the basis of their lists for certain archeological tables. The Israel Museum inaugurated the first multilingual bi-directional museum collections database; supporting fully both Hebrew and English. The Image Search Engine of the Israel Museum, Jerusalem (IMAGINE) is used by curators, restorers, and the registrar's office. The database contains 95,000 object cards (catalogue cards) and 100,000 satellite cards (restoration cards, gallery cards, artist cards). These cards are illustrated with over 20,000 digital images.

2 IMAGINE – The Israel Museum Thesaurus http://www.imj.org.il/imagine/HightLight.asp
12. **Representation of selected controlled vocabularies in RDF/SKOS**

12.1 **Relevant Vocabularies available in RDF/SKOS**

As indicated above some vocabularies relevant for the thematic domain are available already in RDF/SKOS. Here we wish to repeat and indicate the availability of the Virtual Authority File VIAF a project of OCLC and National Libraries. The inclusion of the National Library of Israel among the VIAF partners will assure, in a medium term perspective (end of 2012) a good coverage of names of interest for Jewish cultural heritage content.

Another central vocabulary is the Library of Congress Subject Headings (LCSH), also available in RDF/SKOS. The LCSH is widely used by libraries and Centers of Jewish Studies in leading libraries in the United States. The National Library of Israel (NLI) has adopted it substituting the previous use of Dewey. The Wurtzweiler Library of the Bar Ilan University maintains the Hebrew version of the LCSH and plans are under way for the establishment of a partnership in this matter between the NLI and Bar Ilan.

12.2 **Representing the eJewish.info Thesaurus in RDF/SKOS**

In cooperation with WP3 of Judaica Europeana in a task carried out by NTUA we have expressed the eJewish.info Thesaurus in RDF/SKOS from a previous version made available by the MOSAICA project in XML. This is a multi-lingual thesaurus available in English, Hebrew, French, Spanish, Russian. Its original purpose is the description of Jewish content available in the Internet. Accessible online; each term provides access to the resources it describes. See here.

The eJewish thesaurus includes 3,700 descriptors. In September 2011 these comprised 1,345 subject, 1,315 personal and 1,040 geographical descriptors.

A sample of the converted thesaurus follows:

```xml
<rdf:Description rdf:about="http://ejewish.info/vocab/thes1_760">
  <dcterms:created>12/14/2002 9:51:15 PM</dcterms:created>
  <skos:prefLabel>Yom Kippur</skos:prefLabel>
</rdf:Description>
```

12.3 **Representing the taxonomy - Synoptic Outline of the YIVO Encyclopedia of Jews in Eastern Europe – in RDF/SKOS**

The outline of the YIVO Encyclopedia of Jews in Eastern Europe provides a general view of the conceptual scheme of this encyclopedia. As such it provides taxonomy appropriate for the classification of additional contents related to Jews in Eastern Europe. The encyclopedia being available online it may serve as a resource for enriching the metadata of content resources classified with the help of this taxonomy. Entries are arranged in the conceptual categories:
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

Geographical-Political Units; Social History and Politics; Religion; Language and Literature; Social Organization, Economics and the Professions; Communications Media; Visual and Performing Arts; Everyday Life; History of Study.

For the Online access see: http://www.yivoencyclopedia.org/

The converted taxonomy is available as follows:

http://www.judaica-europeana.eu/skos/YIVO_Encyclopedia_Synoptic.skos (1Mega)
http://www.judaica-europeana.eu/skos/01_geographical.skos (108K)
http://www.judaica-europeana.eu/skos/02_social_history.skos (169K)
http://www.judaica-europeana.eu/skos/03_religion.skos (183K)
http://www.judaica-europeana.eu/skos/04_language.skos (258K)
http://www.judaica-europeana.eu/skos/05_social_organization.skos (91K)
http://www.judaica-europeana.eu/skos/06_communications.skos (93K)
http://www.judaica-europeana.eu/skos/07_visual.skos (62K)
http://www.judaica-europeana.eu/skos/08_everyday.skos (50K)
http://www.judaica-europeana.eu/skos/09_history_study.skos (35K)

Sample:
13. Application of the RDF/SKOS Vocabularies of the Israel Museum Jerusalem for access to Europeana/Judaica Europeana

Following a suggestion by Dr. Allison Kupietzky from the Israel Museum Jerusalem (IMJ) consultations are being held involving staff from Europeana – Antoine Isaac and David Haskiya; Nasos Drosopoulos from NTUA, a partner in Judaica Europeana; and Ram Shimony from the Israel Ministry of Culture which supported the conversion of the controlled vocabularies of the IMJ to SKOS. The purpose is to apply the controlled vocabularies of the IMJ for seeking Jewish related content in Europeana/Judaica Europeana.

The Europeana Search API will be used to build a light-weight app that checks the keyword entered and if it matches an entry in the dictionary adds the other language labels before sending the query to Europeana. (David Haskiya).

**Two ways of browsing IMJ terms:**

or searching through alphabetical listing [http://www.imj.org.il/imagine/thesaurus/objects/objectTOC.htm](http://www.imj.org.il/imagine/thesaurus/objects/objectTOC.htm)

If you use the tree then you open this link: [http://www.imj.org.il/imagine/thesaurus/alobject.htm](http://www.imj.org.il/imagine/thesaurus/alobject.htm)

for example if you like to explore the terminology related to the term **T15505 Items for Jewish events**

press on the + and the tree opens up as such: **T15505 Items for Jewish events**
Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

Press on the + for Jewish items - death and this is what is available:
[+] T15623 Jewish items - death

Burial society token
In order to see the 'behind-the-scenes' script for this term, use this link:
http://www.imj.org.il/imagine/thesaurus/objects/T229673.rdf

in "human" language it looks like this:
http://www.imj.org.il/imagine/thesaurus/objects/T229673.htm

T16115 Collection container
Alternate term: Bag for donations
T16107 Burial jug
T239531 Burial Society cup
Alternate term: Hebra Kadisha cup
T205265 Burial Society glass
Alternate term: Hebra Kadisha glass
T229673 Burial society token
Alternate term: Hebra Kadisha token
T17372 Hebra Kadisha carriage
Alternate term: Carriage, Chevra Kadisha
T16138 Charity box
Alternate term: Money box for zedaka (charity)
Alternate term: Zedaka money box
T222766 Cover for coffin
Alternate term: Coffin cover
T16108 Gravestone
Alternate term: Tombstone
T16111 Hebra Kadisha containers
T16109 Hebra Kadisha cup
T16110 Hebra Kadisha sheet
T16117 Memorial box
T16113 Memorial page
T16136 Money box
T16120 Mourning headscarf
T16121 Shrouds
T16119 Zedaka bag

This list would be a list of search terms which would bring up items that are already present in Europeana that relate to Judaica.
The list that can be seen is for humans so that if you would like to know more about the term (for example):

Burial Society cup
you would press on the link T239531 which would open up
http://www.imj.org.il/imagine/thesaurus/objects/T239531.htm

This is again a human readable format of the broader, narrower and translated terms
For a machine readable format an RDF version has also been published online here:
http://www.imj.org.il/imagine/thesaurus/objects/T239531.rdf
This is the version that will be employed to develop the proposed search mechanism.
14. References

References for the Library Section


References for the Metadata Section


Semantic interoperability report with representation of selected controlled vocabularies in RDF/SKOS

References for the Linked Data Section

References for the Vocabularies Section
McCulloch, E., Macgregor, G. (2008). Analysis of equivalence mapping for terminology services Journal of Information Science 2008; 34; 70 originally published online May 31, 2007; http://jis.sagepub.com/cgi/content/abstract/34/1/70

References for the Section on Places

References for the Section on Periods and Time

Creative Commons License http://www.creativecommons.org
You are free to copy, distribute, display and perform this work under the following conditions: