ECP-2008-DILI-538025

JUDAICA Europeana

Judaica Europeana experiment in Knowledge Management

in partnership with the Haskala Project:

Haskala: Building a modern Jewish Republic of Letters in the 18th and 19th Century

Tel Aviv University, Bar Ilan University and University of Potsdam

Deliverable number: ID2.2
Dissemination level: Public
Delivery date: 1 March 2011
Status: Draft
Author(s): Dov Winer

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

Introduction

The research project on the **Haskala: Building a modern Jewish Republic of Letters in the 18th and 19th Century** which ran from 2000 to 2010 included the development of a comprehensive database concerning the books, authors, subscribers and other aspects related to a substantial corpus of books and other documents from this period. This was a collaborative project headed by Prof. Shmuel Feiner from the Department of Jewish History, University Bar Ilan and Prof. Zohar Shavit, the Unit for Research of the Culture, Tel Aviv University. Prof. Christoph Schulte from the School of Jewish Studies of the University of Potsdam was part of the leading team that initiated this project.

The project established as a main research tool an SQL database for which a users’ interface was developed using Visual Basic. This interface enables the researchers to enter data relating to the different objects and to query the database according to some dozens of pre-set queries. This legacy system prevents the smooth use of the database by the research team. They would like to improve their conditions of collaborative work and to publish substantial parts of the database on the Internet for the benefit of other researchers interested in the Jewish Enlightenment.

**Judaica Europeana** is one of the projects building Europeana, the European Digital Library. It is uploading 5 million digital documents and objects that express Jewish participation in urban life in Europe. Its work plan includes substantial efforts to have such documents used in a variety of contexts: genealogical research, education, virtual exhibitions, cultural tourism, university teaching and scholarly research. The work plan includes running an experiment of collaborative work among participants of a community of knowledge e.g. a group of scholars working together on research focused on a given corpus.

These complementary interests of the Haskala project and Judaica Europeana are the motivation for the following proposal.

**The Haskala Republic of Letters Research Project**

This project seeks to investigate the secularization of the traditional book culture and consider the mechanisms involved in this process. In particular it focuses on researching a public of producers and readers and the channels that supported the dialogue among them. In the first years of the Haskala the public involved in its activities was small. To enlarge the scope of its reach the Maskilim employed different means to market their products and mobilize different communities. They established schools to develop a young Maskil generation; they established associations of Maskilim; founded printing houses or made agreements with existing ones; established periodicals and authored books to disseminate their ideas. From these books gradually disappeared the traditional rabbinic “Askama” (Approval) as a lever for the marketing of the book and its evaluation. The printers of the books, the producers, or the authors themselves promised to the readers that they will benefit from the books in a way that is consonant with a modern secular ethos. The readers were approached as intellectuals in the modern sense of this concept and not on the grounds of the religious authority of the author of the text.

This research is based on a comprehensive database about the books that were published in the German-speaking space as part of the Haskala movement. The software supports 147 queries about different characteristics of the books, their authors and the readers. The database includes hundreds of texts in Hebrew and German that were published in the German speaking space. It supports the kind of...
Knowledge Community

analysis required to study the range of activities and the context of the Haskala
Republic and describe hundreds of agents that operated as subscribers, financers,
authors and printers.
For further details see (Hebrew):  http://www.tau.ac.il/tarbut/zohar.shavit/ZS-
Proyeqtim2007-2008.htm

Judaica Europeana: supporting semantic assisted collaborative knowledge
communities
The large digital corpus of documents that Judaica Europeana is establishing should
be employed in the creation of new knowledge. This can be achieved by supporting the
work of scholars concerned with the participation of Jews in the development of the
European urban space and related issues. Such support is related to the application of
the tools developed in the area of Digital Humanities scholarship. Judaica Europeana
participated and initiated several events in view to advance this purpose. The present
paper seeks to define the parameters for an exemplary pilot experiment.
Digital Humanities seek to integrate technology into scholarly activities. It is defined
methodologically by the belief that means of knowledge-making, dispersal, and
collection are common among the disciplines that make up the liberal arts. John
Unsworth defines these common activities as: discovering, annotating, comparing,
referring, sampling, illustrating and representing.
Digital Scholarship is defined in the report of the American Council of Learned
Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences
(Our Cultural Commonwealth, 2006) as follows:
a) Building a digital collection of information for further study and analysis
b) Creating appropriate tools for collection-building
c) Creating appropriate tools for the analysis and study of collections
d) Using digital collections and analytical tools to generate new intellectual products
e) Creating authoring tools for these new intellectual products, either in traditional
forms or in digital form. They note that all of this requires a great deal of
cooperation; it is still imaginable that (d) can be the work of a single individual.

One of the International Advisors to that Commission, Prof. Stefan Gradmann,
presented his vision how Europeana will support Digital Humanities, in his key-note
lecture “Europeana Semantica”. The following figure provides a summary of that
vision. It is the process that may lead from the digital corpora/sources, through their
modeling including the application of tools for semantic analysis, and continuing with
the interpretation and collaborative work that result in scholarly publications
expressing new insights and knowledge.

---

1 John Unsworth (2005) Scholarly Primitives: What methods do humanities researchers have in common, and how might
our tools reflect this? http://jefferson.village.virginia.edu/~jmu2m/Kings.5-00/primitives.html

2 Our Cultural Commonwealth, report of the American Council of Learned Societies Commission on
Cyberinfrastructure for the Humanities and Social Sciences (2006)
http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf

3 Stefan Gradmann (2008) Europeana Semantica remarks on the motivations for our semantic foundations, Key note
lecture at the 5th EVA/MINERVA Conference, Jerusalem, November 2008:
http://www.slideshare.net/gradmans/europeana-semantica
Europeana has made big advances in developing its capabilities to support this process. It defined a new data model (EDM) based on the principles and tools of the new structured semantic Web of Linked Data. EDM not only supports the expression of the full richness of the content providers’ metadata, it also enables data enrichment with contextual information from a range of third party sources. The main standards employed in EDM are OAI-ORE for organization of metadata about an object and SKOS for vocabulary representation.

The goal of the Linking Open Data community 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 APIs.

The Linked Data approach emphasizes the re-use and linkage of richly described resources on the web. This is consonant with 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’ spaces, or to external spaces used as knowledge references.

---

1 See EDM Data Model Primer in [http://group.europeana.eu/web/europeana-project/technicaldocuments/](http://group.europeana.eu/web/europeana-project/technicaldocuments/)


3 SKOS Simple Knowledge Organization System Primer [http://www.w3.org/TR/skos-primer/](http://www.w3.org/TR/skos-primer/)

4 Linked Data [http://linkeddata.org](http://linkeddata.org)

5 Linking Open Data [http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description](http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description)

6 Europeana Data Model Primer 05/08/2010 [http://tinyurl.com/edmprimer](http://tinyurl.com/edmprimer)
Judaica Europeana prepared a document that provides basic explanations of the concepts involved in the structured semantic Web. The document details a program of work required to identify those vocabularies, related to Jewish history and culture, which are necessary for exploiting the potential of this new environment for Jewish related scholarship. The document focus is on vocabularies concerning Names (Who?), Places (Where?) and Periods (When?).

**Collaboration**

Digital Humanities research has been grounded in the development of a variety of tools enabling scholars to collaborate when they analyze, annotate, share and publish works based on digital corpora. A non-exhaustive compilation of initiatives and tools was prepared by Judaica Europeana for its workshop on Digital Humanities held in Ravenna in July 2010 (see [http://www.judaica-europeana.eu/digital-resources.html](http://www.judaica-europeana.eu/digital-resources.html)). Klamma et al. (2005) present an early summary of projects supporting collaborative scholarly work. He grounds it on a transcriptive theory for discourses in the Humanities. Successful communication and knowledge organization depends on the choice of appropriate media for specific scientific discourses. They refer, as an example, how discourse knowledge was encoded within Talmudic tractates. Their transcription into a structured and annotated multi-lingual hypertext supports the switching between languages (English, German, Hebrew) while maintaining the discourse structure through color highlighting and annotations. Such features make these texts accessible to students and other interested scholars with limited knowledge in both Hebrew and Judaist concepts. This re-addressing has rapidly created a worldwide teaching and learning community.

Communities of practice (CoP) of scholars are characterized by common conventions, language, tool usages, values and standards (Wenger, 1998 cited in Klamma et al, 2005). The development of a common practice which defines the community comprises the negotiation of meaning among the participants as well as the mutual engagement in joint enterprises and a shared repertoire of activities, symbols and artifacts. A CoP is inseparable from issues of (individual and social) identity. Identity is mainly determined by negotiated experience of one’s self in terms of participation in a community and the learning process concerning one’s membership in a CoP. Systems supporting CoP should aim at providing scholars with a flexible (online) environment to create, annotate and share media-rich documents for the discourses by relying on metadata standards. Those standards allow scholars to create, exchange and collaborate on multimedia artifacts and collections between communities across disciplines and distances (Klamma et al., 2005). For the purpose of comparing different approaches in developing systems that support CoP, Klamma details the requirements for such systems. These include: (1) **Collective hypermedia artifact repository** (2) **Transcription and semantic enrichment of data**. Collaborative learning


2 Digital humanities resources (tools, reports, centers, conferences, research) [http://www.judaica-europeana.eu/digital-resources.html](http://www.judaica-europeana.eu/digital-resources.html)

3 The Judaica Europeana Digital Humanities Workshop , University of Bologna, Ravenna Campus sponsored by COST Action 32 Open Scholarly Communities on the Web (July 2010) [http://www.judaica-europeana.eu/events.html](http://www.judaica-europeana.eu/events.html)
is encouraged by annotations accessible to and possibly transcribed by other scholars. (3) **Search and retrieval** – all the processes of retrieval, manipulation and management should be accessible as objects in the repository. (4) **Community management** – flexible, providing with access rights on different levels and for different roles. (5) **Personal and group collections.** The need for fluid archives on personal and group level, which can be navigated, sorted and annotated by community members. (6) **Hypermedia and interrelation graphs.** The expression of the full context and complexity of objects requires the visual representation of knowledge by hypermedia graphs. (7) **Ontologies.** These are applied for information brokering and provide users with content deemed the most suitable in a particular context. Klamma points out the problems and limitations of ontologies.

**Annotation**

Jane Hunter (2009) has carried out a comprehensive review of collaborative semantic tagging and annotation systems. Annotating is used to organize, create and share knowledge. Individual scholars use it when reading, as an aid to memory, to add commentary, and to classify documents. It can facilitate shared editing, scholarly collaboration and pedagogy. She points out that many of these tools are designed for specific collection types, user requirements, disciplinary application or individual desktop use. Scholars are confronted with having to learn different annotation clients for different content repositories, have no easy way to integrate annotations made on different systems and/or created by colleagues using other tools, and are often limited to simplistic and constrained models of annotation. For example, many tools support the simplistic model in which the annotation content comprises a brief unformatted piece of text; may tools conflate the storage of annotations and the target being annotated.

(Hunter et al. 2010) describes the Open Annotation Collaboration (OAC), a data model to support sharing and interoperability of scholarly annotations. In the OAC model an Annotation is an Event initiated at a date/time by an author.

**The development of the present proposal – acknowledgments**

Since its inception in July 2010 the proposed configuration of the present proposal has several origins: (1) A continuing monitoring of recent developments through the literature (2) The beginning of implementation of Europeana new data model, EDM. (3) A dialogue with some leading research and development teams: COST A32 – Open Scholarly Communities on the Web\(^1\). This action, concluded at the end of 2010, sought to create a digital infrastructure for collaborative humanities research on the Web and to establish and foster the growth of Scholarly Communities. It served as an incubator for many initiatives and clustered several initiatives using the tools developed in the Discovery project\(^2\). The tools they developed include **Talia** which is an open-source, web-based, distributed digital library and publishing system, designed specifically for humanities research and Philospace\(^3\). The main developers are Michele Barbera and Christian Morbidoni and they were among the main

---

2. Discovery Project: Philosophy in the Digital Era – read, study, cite and publish in a semantic web environment [http://www.discovery-project.eu](http://www.discovery-project.eu)
3. Discovery Project Technologies [http://www.discovery-project.eu/technologies.html](http://www.discovery-project.eu/technologies.html)
presenters in the Judaica Europeana Digital Humanities workshop held in Ravenna. This program of research and development is being pursued by the startup Netseven, and by Muruca, a collection of Open Source applications to create, manage and run Semantic Digital Libraries.

The laboratory STIH "Sens, Texte, Informatique, Histoire" at the University of Paris-Sorbonne includes the team lead by Prof. Philippe Laublet with Milan Stankovic and Alexandre Passant. The tools they developed include MOAT (Meaning of a Tag) and the SemSlates framework. The presentation by Prof. Laublet and Milan Stankovic at the Judaica Europeana meeting held in February 2011 considerably advanced the crystallization of the present document.

Finally, following the suggestions included in the presentation by Laublet and Stankovic we got in touch with Yaron Koren, one of the main developers of the Semantic Media Wiki.

Semantic MediaWiki (SMW) is a free, open-source extension MediaWiki – the wiki software that powers Wikipedia – that allows the storing and querying of data within the wiki’s pages. Semantic MediaWiki is also a fully-fledged framework, in conjunction with many spinoff extensions, which can turn a wiki into a powerful and flexible “collaborative database”. All data created within SMW can easily be published via the Semantic Web, allowing other systems to use this data seamlessly. Following these developments, we have decided to use the Semantic MediaWiki (SMW) as the core tool for the present pilot.

Semantic Media Wiki as the core tool for the project

Semantic MediaWiki has grown a long way from its roots as an academic research project. It is currently in active use in hundreds of sites, in many languages, around the world, including Fortune 500 companies, biomedical projects, government agencies and consumer directories. The Wikipedia article on Semantic MediaWiki contains a section listing some of its notable usages. You can also see one list of the sites that use SMW here, and another here. It should be noted that both lists focus on public sites, although perhaps half or more of the sites that use Semantic MediaWiki are private, for internal use by companies and organizations. It is in use on over 200 public active wikis around the world, in addition to an unknown number of private wikis.

---

1. [http://www.judaica-europeana.eu/events.html](http://www.judaica-europeana.eu/events.html)
2. Netseven [http://www.netseven.it](http://www.netseven.it)
6. Alexandre Passant [http://www.deri.ie/about/team/member/alexandre_passant/](http://www.deri.ie/about/team/member/alexandre_passant/)
There are a number of consulting companies that implement SMW as part of their solutions, including Benchmarking Partners, FZI, LeveragePoint, ontoprise and WikiWorks.

SMW has noticeably gained traction in the health care domain for collaboratively creating bio-medical terminologies and ontologies. Examples are LexWiki, which is jointly run by the Mayo Clinic, National Cancer Institute (NCI), World Health Organization (WHO) and Stanford University; Neuroscience Information Framework’s NeuroLex; and the Concept Hub wiki, operated by 3M Health Information System.

**Some examples of Semantic Media Wiki Applications (SMW of the Month)**
The following are some examples of applications of SMW that may provide a glimpse of the expected results of its implementation for the Haskala project:

Institute of Applied Informatics and Formal Description Methods (AIFB)
http://semantic-mediawiki.org/wiki/AIFB_Web_Portal

**Familypedia** is a free-to-use public website for genealogy and family history
http://semantic-mediawiki.org/wiki/Familypedia

**Our Brant** is a local-history wiki created by the County of Brant Public Library
http://semantic-mediawiki.org/wiki/Our_Brant

**Semantic Media Wiki with a Linked Data Integration Framework**
At an event held last February, Christian Becker, one of the authors of DBPedia breakthrough project of Linked Data presented an application of the Semantic Media Wiki with a Linked Data Integration Framework. This project is of particular interest for our proposed pilot. The Europeana new data model

The functionalities that will become available
Following the intervention proposed in this the following facilities will become available. They will support the work of the original community of scholars who build
the database. It will also serve a larger community of scholars and students with the publication of selected parts of the database in Europeana and as Linked Data.

- **Improved data structure.** In place of categories for structuring data, simple queries will reduce the need for a complex classification system. Semantic templates enable the storage of semantic markup the wiki will further develop its a solid data structure. The Semantic Forms extension lets administrators create forms for adding and editing the data within semantic templates, thus making the addition of semantic information possibly even easier and more straightforward than regular wiki text.

- **Searching information.** Individual users can search for specific information by creating their own queries reducing the dependences of the researchers on the developers.

- **Automatically-generated lists.** In SMW, lists are generated automatically like this. They are always up-to-date and can easily be customised to obtain further information.

- **Visual display of information.** The various display formats defined by additional extensions, such as Semantic Result Formats and Semantic Maps, allow for displaying of information in calendars, timelines, graphs and maps, among others, providing a much richer view of the data than simple lists would.

- **Inter-language consistency.** In wikis that span multiple languages, like Wikipedia, there is often a great deal of data redundancy, which can lead to inconsistencies. For example, the population of Edinburgh *at the time of this writing* is different in the English, German, and French Wikipedias. If data is stored semantically, you could, for instance, ask for the population of Beijing that is given in the Chinese Wikipedia without knowing a single word of that language. This can be exploited to have different languages query one another's data, either for reuse or at least to detect inconsistencies.

- **External reuse.** Data, once it is created in an SMW wiki, does not have to remain within the wiki: it can easily be exported via formats like CSV, JSON and RDF. This enables an SMW wiki to serve as a data source for other applications, or, in the case of enterprise usages, to take over the role that a relational database would normally play. Through the use of the External Data extension and the result format Exhibit, one SMW-based wiki can even use the data from another, eliminating the need for redundancy between wikis. You can also query SMW's data via an RDF triplestore, using any of the available triplestore connector extensions.

---

Integrate and mash-up data. Data contained in an SMW installation does not have to be an isolated store of information. Extensions such as the Data Import, Data Transfer and External Data extensions empower you to integrate external data (coming e.g. from legacy systems, web services or linked data sources) and interrelate it with existing semantic data in the wiki. Thus, an SMW-powered wiki can serve as a central information hub in an IT landscape.

Main Tasks/phases for the proposed pilot project

Annex 1 describes the main tables of the present database of the Haskala project. A list of the supported queries should supplement this document. See: Annex_1_Haskala_Main_Tables.doc

The following tasks are proposed as part of a pilot project:

1. Conversion of the database to a RDF store and its publication both as RDF and HTML.

   The conversion of the database can be carried out either through the D2R suite\(^1\) or using the native import features of the Semantic Media Wiki. The later is able to digest either XML or CSV formats.

   The D2R server uses the D2RQ mapping language to capture mappings between specific database schemas and RDFS schemas. Such mapping specifies how resources are identified and how property values are generated from database content. The central object in D2RQ is the ClassMap that represents a mapping from a set of entities described within the database, to a class or a group of similar classes of resources. D2R Server includes a tool that automatically generates a D2RQ mapping from the table structure of a database. The tool generate a new RDF vocabulary for each database, using table names as class names and column names as property names. The mapping can be customized afterwards by substituting auto-generated terms with terms from well-known RDF vocabularies.

   Alternatively the Data Transfer extension\(^2\) to MediaWiki can be employed. It allows users to both export and import data from and to the wiki, with export done in XML format and import possible in both XML and CSV formats.

2. Enrichment of the resulting RDF store through interventions like:

   a. Inclusion of links to available digitised versions of the books described in the database and to their catalog entries in WorldCat.

   b. Whenever available substitution of keywords included in the database as textual strings by their URI in qualifying vocabularies.

   c. A good example is the substitution of the Names (of authors and other actors) by the corresponding URI from the VIAF vocabulary (available in SKOS). See:

---


\(^2\) SMW Data Transfer extension http://www.mediawiki.org/wiki/Extension:Data_Transfer
http://viaf.org/
http://www.oclc.org/developer/services/viaf

d. Whenever available other relevant vocabularies (e.g. for places) will be used as well for this process of enrichment.

3. Publication of the database as LOD and upload to Europeana

See the new Europeana Datamodel (EDM):
EDM Data Model Primer  EDM Data Model Definition v5.2

4. Querying the database

Querying facilities should be established enabling the researchers to retrieve data from the RDF store following their specific requirements. This can be set using the native facilities of the SMW and additionally defining SPARQL queries to be carried out over appropriate end points of LOD. On March 2011 Europeana published its API and a Widget. These can be integrated in the proposed system providing additional enrichment to the query process.

5. User interface and collaborative facilities
   
a. A user interface enabling the update of data the RDF store so that the use of the legacy database can be discontinued

6. Hebrew: several tables of the Haskala database are in Hebrew. We point out here that the SMW has full Hebrew support and this is a distinct advantage among the considerations that lead us to adopt such solution.
## Table of tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Responsible</th>
<th>Supported by</th>
<th>Observations</th>
</tr>
</thead>
</table>
| Agreement between the Haskala project: University of Frankfurt and Judaica Europeana | Rachel Heuberger | Dov Winer assisted by Lena Stanley-Clamp | 1. Prototype  
2. Estimate of resources  
3. Define roles of parties to the agreement  
After first tasks have been carried out |
| Semantic Media Wiki Installation                                     | Marko Knepper | Yaron Koren                        |                                                                              |
| Database export                                                      | Ayala Hildesheimer | Yaron Koren                        | Export either as CSV or XML files                                             |
| Database import into SMW                                             | Marko Knepper | Yaron Koren                        |                                                                              |
| Database enrichment                                                  | Rachel Heuberger | National Library of Israel          | - Identifying existing digitised copies of the books described in the database and substituting in SMW  
- Retrieving books entries in the WorldCat and including them in the SMW |
| Metadata enrichment                                                  | Main responsibility for leading this process should be clearly allocated to a staff person either from the Haskala project or the University of Frankfurt | -Esther Guggenheim VIAF maintainer at the NLI Europeana team  
-Terminalogy experts (e.g. Iris Blochel-Dittrit)  
-Dov Winer | Identifying vocabularies in the Haskala DB that can be mapped into concepts of authorities: Names, Places, Periods VIAF, Geographical Names, Getty etc |
<p>| Importing the existing                                               | Yaron Koren | Marko Knepper                      |                                                                              |</p>
<table>
<thead>
<tr>
<th>Task</th>
<th>Responsible Parties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haskala DB queries into SMW queries</td>
<td></td>
</tr>
<tr>
<td>Designing the new SMW work environment for the Haskala research team</td>
<td>Yaron Koren</td>
</tr>
<tr>
<td>Implementing the new editing and display facilities</td>
<td>Yaron Koren, Marko Knepper</td>
</tr>
<tr>
<td>Defining groups of administrators and users their level of access and providing authorizations</td>
<td>Marko Knepper, Yaron Koren</td>
</tr>
<tr>
<td>Graphic design</td>
<td>University of Frankfurt, Haskala research</td>
</tr>
<tr>
<td>Publishing part of the database [defined by the Haskala research team] as Linked Data</td>
<td>Marko Knepper, Yaron Koren</td>
</tr>
<tr>
<td>Publishing part [defined by the Haskala research team] of the database in Judaica Europeana</td>
<td>Marko Knepper, Stefanos Kolias, Nasos</td>
</tr>
<tr>
<td>Training of the Haskala research team and their research assistants</td>
<td>Yaron Koren</td>
</tr>
<tr>
<td>Routine use and maintenance with a perspective of several years</td>
<td>University of Frankfurt, Haskala Research</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some of the training may be carried out in a face to face meeting in Israel. The possible preparation of a Webinar enabling participants from different countries should be considered. Guidelines should be established between these two partners for the project.