# Mapping Metadata for SWHi: Aligning Schemas with Library Metadata for a Historical Ontology

Junte Zhang*, Ismail Fahmi, Henk Ellermann, and Gosse Bouma

Humanities Computing Department and University Library
University of Groningen
Oude Kijk in't Jatstraat 26, 9712 EK Groningen, The Netherlands

**Abstract.** What are the possibilities of Semantic Web technologies for organizations which traditionally have lots of structured data, such as metadata, available? A library is such a particular organization. We mapped a digital library's descriptive (bibliographic) metadata for a large historical document collection encoded in MARC21 to a historical ontology using an out-of-the-box ontology, existing topic hierarchies on the World Wide Web and other resources. We also created and explored useful relations for such an ontology. We show that mapping the metadata to an ontology adds information and makes the existing information more easily accessible for users. The paper discusses various issues that arose during the mapping process. The result of mapping metadata to RDF/OWL is a populated ontology, ready to be deployed.

## 1 Introduction

The Early American Imprints Series I [1] are a microfiche collection of all known existing books, pamphlets and periodical publications printed in the United States from 1639-1800, and gives insights in many aspects of life in 17th and 18th century America, and are based on Charles Evans' American Bibliography. This bibliography has been created in MARC21. Identifying and characterizing a resource and placing it in an intellectual context is expensive. The 'expensive' metadata are not fully used by the library's users, and hence the resources are not fully disclosed using the existing metadata as an extra supporting layer.

This paper will present a method to make these existing bibliographic (descriptive) metadata more easily accessible to (casual) users using Semantic Web technologies. The Semantic Web builds on information that is machine-readable and allows links to be created with relationship values [1]. Ontologies are the backbone of this idea, because these are used to organize and store information. Ontologies are built independently of a given application, and ensure that there is a common understanding of a domain (interoperability) [3].

There is related work in other domains, where thesauri such as the Arts and Architecture Thesaurus (AAT) are being used to create an ontology [8], and there

* Current affiliation is Archives and Information Studies, University of Amsterdam, email: j.zhang@uva.nl

1 `http://library.truman.edu/microforms/early_american_imprints.htm`

is more merging of datasets and vocabularies in the Cultural Heritage domain in [6]. Similar research has been conducted by [7] where a medical thesaurus called Medical Subject Headings (MeSH) and WordNet are converted to RDF/OWL.

However, these papers do not specifically deal with digital libraries' ample bibliographic metadata formatted in MARC21, which is often 'noisy', and what characteristic issues arose while aligning MARC21 metadata with schemas or vocabularies and its mapping process. Much metadata of libraries or archival institutions is encoded in this format, and other organizations with semi-structured data may face similar challenges. We also wondered whether the idea of mapping metadata, storing it in repositories using Semantic Web techniques, and advanced querying with inferencing, is feasible for our library or others, and could really be useful for people interested in the history of the United States.

This research was carried out as part of the *Semantic Web for History* (SWHi) project, of which a system description is presented in [2].[1] An objective of this project is to explore how a historical ontology can be built using library metadata, allowing libraries to push the Semantic Web forward, with real use for historians and other users. For example, the system should be able to answer questions such as:

1. When was George Washington born, and when did he die?
2. What events have occurred in the Early American History?
3. What did George Washington publish?

These questions retrieve factoid answers, and the three questions could be synthesized in this form:

4 Did George Washington publish about the events that have only occurred in his life?

On the one hand, we wonder whether the populated historical ontology can answer the first 3 questions, and use our historical ontology to infer the answer of the fourth question. On the other hand, an objective is to find out whether an ontology can offer more focused access to a specific nugget of information that captures the user's information need. The methodology of our approach is explained in Section 2. Preliminary results are presented in Section 3. We conclude with our findings and point to future work in Section 4.

## 2 Methodology

### 2.1 Data Exploration

Our bibliographic MARC21 metadata have been created by librarians. To know what is in the data, and to know what information is needed, the MARC21[2] file was analyzed in detail. We started by encoding this file in XML. A MARC21

[1] `http://semweb.ub.rug.nl/`
[2] `http://www.loc.gov/standards/marcxml/`

record encoded in XML consists of a set of datafield blocks, and each of these datafields has got numerous subfields as attributes, which makes MARC21 an expressive and rich metadata standard. The combination of a datafield and a subfield indicates the semantics of the metadata value. In this example, the datafield is *100* and the subfield code is *a*, which means the name of an author.

```
<datafield tag="100" ind1="1" ind2="">
   <subfield code="a">Gardiner, John Sylvester John,
   </subfield>
   <subfield code="d">1765-1830.         </subfield>
</datafield>
```

The metadata of the Evans bibliography (150 Mb, 36,305 records) are the single source of input for populating the ontology. There are 772,258 datafield items, classified in 35 types (using the TAG attribute). There are 1,647,280 subfield items (of which almost 40% is subfield *a*), classified in 27 different types (using the CODE attribute). If we also take the type of the datafield into account, there are 190 different combinations of datafield type/subfield type pairs. The most frequent combinations are given in Table 1.

**Table 1.** Snippet of the mapping table.

| **MARC21** | # | % | $\sum$ | **Description** | **Schema** | **Used** |
|---|---|---|---|---|---|---|
| 035 a | 72610 | 4.41 | 4.41 | SYS. CONTROL NO | N/A | N/A |
| 510 c | 69262 | 4.20 | 8.61 | CITATION | dc:relation | Y |
| 510 a | 69262 | 4.20 | 12.82 | CITATION | dc:relation | Y |
| 500 a | 68244 | 4.14 | 16.96 | GENERAL NOTE | dc:description | Y |
| ... | ... | ... | ... | ... | ... | ... |
| 600 k | 1 | 0.00 | 100.00 | PERSONAL NAME | | N |

Note that the metadata consists of mappable *descriptive* (bibliographic), *structural*, and *administrative* metadata. The latter 2 are not mapped, because it does not have meaning in the historical domain of the ontology, and it is not useful for (non-librarian) users. We give a quantitative overview of the portion of metadata that was mapped, and discuss the contribution of each schema.

### 2.2 Reusing Existing Resources

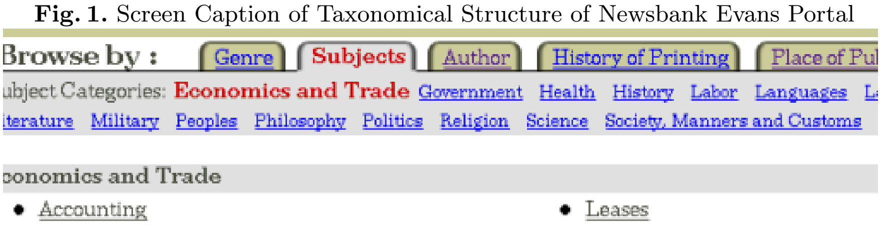
We do not convert the MARC21 metadata directly one-to-one to RDF, which would be trivial, but try to use extra knowledge, links and descriptives provided by different resources and align them together comprehensively. A plethora of existing resources can be reused and merged to create a single historical ontology. We have decided to use an existing historical ontology as the base, and modify and enrich it with existing resources. By reusing existing knowledge structures, there is greater acceptance for the ontology. Prefixes and namespaces are essential to align schemas and map ontologies, because it indicates where an instance (individual), class (concept) or relation is derived from. It is also a matter of accountability and accuracy, which is important for a historical ontology.

The new ontology SWHi $\Omega$ can be seen as the union of different subsets or $\Omega = \{V \cup D \cup S \cup F \cup N\}$, where $V$ stands for the VICODI core ontology[3], $D$ for the Dublin Core predicates for describing *documents*[4], $F$ for the FOAF predicates and classes for describing the *social networks*[5], $N$ for the Newsbank Topic Hierarchy (taxonomy) classes for *topic classification*[6], and $S$ for the SWHi predicates and classes. An instance $i$, where $i \in \Omega$, is described by combining the predicates from these different subsets.

**VICODI** A history-specific ontology was built by [4], because there were no suitable existing ontologies available. The VICODI structure is intuitive, simple and allows for the uploading of instances and relations (representing historical facts) into the ontology in large numbers. Although VICODI is intended for European history, it can also be used for American history. Any shortcomings can be dealt with by modifying or enriching the ontology.[7]

**Newsbank Topic Hierarchy** (NTH) We have substituted VICODI's Category hierarchy with new classes based on the taxonomical structure of the webpage of NewsBank's Readex archive of the Evans dataset. This taxonomical structure was ready made and is depicted in Fig. 1.

**Fig. 1.** Screen Caption of Taxonomical Structure of Newsbank Evans Portal




All the subclasses of the tab *Subjects* are manually extracted as the categories of the Imprints. 16 categories in total are extracted from the NTH. Each of the categories has numerous (1909) topics listed. These are saved as HTML files and semi-automatically fetched and fed as instances to each of the 16 corresponding

[3] `http://www.vicodi.org/about.htm`
[4] `http://dublincore.org`
[5] `http://www.foaf-project.org`
[6] `http://infoweb.newsbank.com`
[7] The latest beta version of the SWHi ontology uses PROTON (`http://proton.semanticweb.org/`) as its core, but the same kind of relations as in VICODI are used as discussed in this paper.

categories. For example, the subject *Accounting* is an instance of class *Economics and Trade*. The subjects are mentioned in the documents (imprints) as topics, so each document can be related to the NTH.

**Type of Imprints** In the metadata, each MARC21 record has the properties of an imprint. In many imprints, it is known what kind of imprint it is. This piece of code shows that an imprint can both be categorized as *Broadsides* and *Hymns* with datafield `655` and subfield `a`.

```
<datafield tag="655" ind1="" ind2="7">
   <subfield code="a">Broadsides.</subfield>
   <subfield code="2">rbgenr      </subfield>
</datafield>
<datafield tag="655" ind1="" ind2="7">
   <subfield code="a">Hymns.</subfield>
   <subfield code="2">rbgenr     </subfield>
</datafield>
```

This structure was not present in the VICODI ontology, so it will be automatically extended with these. 131 types of imprints are extracted from the metadata and used as classes for the ontology.

**Dublin Core** (DC) It is an annotation vocabulary for metadata. There is a distinction between a qualified and unqualified (or simple) version, because the former has been intended for finer semantic distinctions and more extensibility, while the latter is simple and concise. Unqualified DC contains 15 elements. Qualified DC uses qualifiers to narrow the scope of an element, e.g. *dc:date.created* is more refined than *dc.date* alone. The 'Dumb-Down Principle' is applicable here, because values of qualified DC can always be mapped to unqualified DC. That is why we have decided to use qualified DC wherever possible.

**Friend of a Friend** (FOAF) This vocabulary is used for describing social networks. It is a suitable schema, because predefined elements of persons or organizations exist in the metadata and can be mapped to RDF/OWL using the FOAF vocabulary, such as names of authors or publishers. FOAF-properties have FOAF-classes as their domain (or range), so FOAF-classes are added to the ontology in order to use FOAF properties.

### 2.3 Adding Class Hierarchies

- **Semi-automatic**. The classes, properties and their subclass relations are defined and stored on top of the OWL file, which is illustrated in this code.

```
<owl:Class rdf:about="http://semweb.ub.rug.nl/newsbank/Science">
  <rdfs:subClassOf rdf:resource="http://vicodi.org/ontology#Category"/>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    newsbank:Science</rdfs:label>
</owl:Class>
```

*Science* is made a subclass of *Category*. We automatically added all 131 subclasses of the *Type of Imprints* to the class *foaf:Document*.
- **Manual**. Protégé[8] arguably has become the most widely used ontology editor for the Semantic Web [5]. We have used it to define classes and their hierarchies, relationships between classes, and properties of these relations. The 16 NTH classes have been manually added, as previously discussed. FOAF-classes and properties have also been added manually using Protégé. The properties of FOAF have FOAF-classes as their domain. *foaf:Group*, *foaf:Organization*, and *foaf:Person* are subclasses of *foaf:Agent*.

### 2.4 Properties and Relations

Schemas are used to add properties (predicates) to existing classes, and to create relations between the classes. These properties have been manually added using Protégé and are saved as RDF/OWL. We have created relations between classes, besides the *subClassOf* relations in RDF. Inference is important: all classes inherit the properties (attributes or slots) of the superclass. RDF resources can either be a literal or another resource (object). In the latter case it is of type *instance* or type *class*. The properties that we have used to create such relations, and enrich the historical ontology, are depicted in Fig. 2. The depicted ISA-relations are equivalent to *subClassOf* relations.

1. **Time Properties** make any object in the ontology temporal:
   - *vicodi:exists*, which contains general descriptions about time,
   - *dc:coverage.temporal* is used for a document to describe the timeframe the imprint is covering (e.g. American Revolutionary War),
   - and *dc:date.publication* which describes when a document was published.

   The domain of *dc:coverage.temporal* is the *vicodi:Time-Dependent* class, and with inheritance, all its subclasses have this property as well, and it refers to *swhi:Ontology*, so it can take any instance in the ontology as value. It does not link directly to the instances of *vicodi:Time*, because this property should also be allowed to take as object instances of *vicodi:Event*, which is a different 'leaf' in the subclass hierarchy of the ontology. All other resources can be made temporal with *vicodi:Time-Dependent*.
2. **Agent Properties** are related to persons or institutions, and described with
   - *dc:creator*, which refers to an agent that created an object (document),
   - *dc:publisher*, which refers refers to an agent of type Publisher,
   - and *foaf:knows* which relates a person to another agent.

   It is assumed that an author expressed with *100 a* knows the names of the persons that he or she has covered in his publications, e.g. with the datafields *600* and *700* in the same MARC21 record. This implicit knowledge in the MARC21 data is made explicit by mapping it using the *foaf:knows* property. Moreover, incoming and outgoing links can be detected for persons.
3. **Topic Properties** classify and cluster objects in the ontology:

[8] http://protege.stanford.edu/

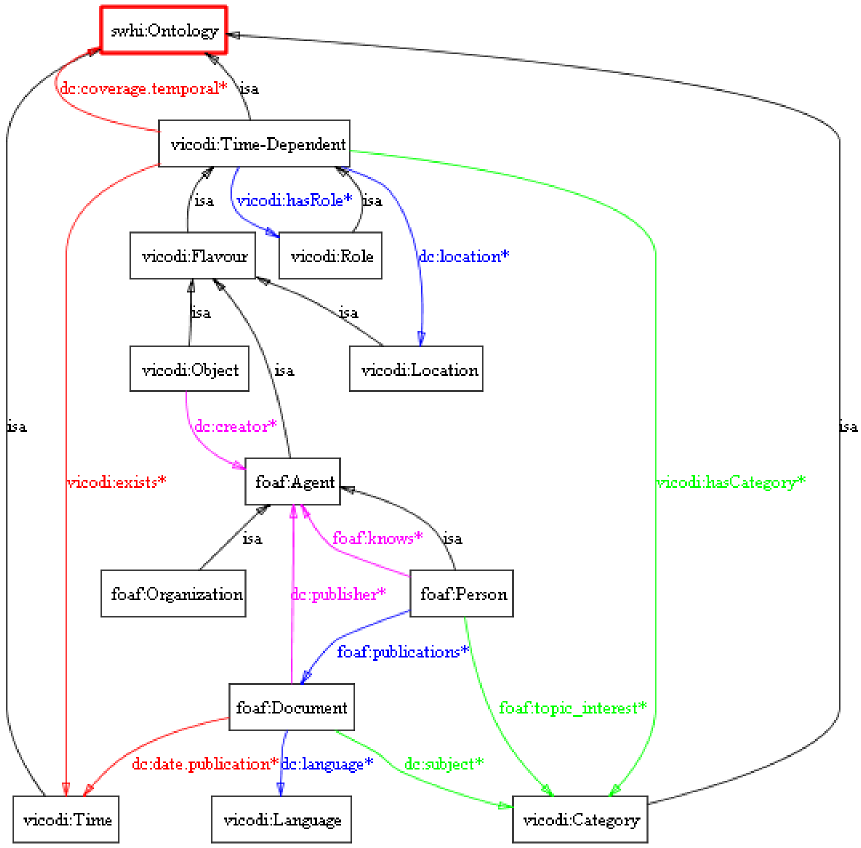



**Fig. 2.** Some classes and key relations in the SWHI Ontology

- *vicodi:hasCategory*, which makes clear that anything in the historical ontology can have a topic, and hence be classified with the *NTH*.
- *foaf:topic_interest*, which specifically refers to the topics that a person is interested in, because he or she has published about it,
- and *dc:subject* which point to the topics from the document collection.

For example, this makes it possible to retrieve all persons who are also interested in a certain topic of an imprint, which could give a user a list of possibly interesting authors as query expansion.

4. **Remaining Properties** are:
   - *vicodi:hasRole*, which makes temporal objects (like persons) having roles,
   - *dc:location* links an object besides time also with space,
   - *foaf:publications* gives a list of publications for each author pointing to the instances of *foaf:Document* and its subclass hierarchy,
   - and *dc:language* makes clear which language is used.

   The *vicodi:hasRole* relation takes a class as its range, as it refers to subclasses (e.g. *vicodi:Author* and *vicodi:Publisher*) of the class *vicodi:Role*.

### 2.5 Automatically Populating the Ontology

This section presents what steps were needed to use the library metadata for automatically populating the ontology and the reasoning behind it.

**Process and Cleanup Data** The metadata was already encoded in XML using MARC21 elements, and thus we could have used XSLT for the conversion. However, as the conversion required a substantial amount of *string processing*, *regular expressions* and *text normalization*, we opted for Perl. String processing is useful a.o. for dissecting temporal intervals and for cleaning up the noisy metadata. Instances are automatically extracted and printed as one line. Duplicates are removed by retrieving only unique lines, after removal of leading and trailing whitespace, squeezing of multiple spaces, and removal of some punctuation.

**Mapping Metadata** The process of mapping the MARC21 to RDF/OWL is done by checking out Table 1 from top to bottom, which is sorted by frequency (descending). The purpose was to reuse multiple existing schemas for mapping knowledge in the metadata, in this case DC, FOAF and VICODI, as much as possible, because existing schemas are accepted by other people and widely used. Another reason is for instance that Dublin Core is not expressive enough to capture the semantics of MARC21. RDF allows us to create our customized SWHi schema as the fourth schema. Four ways to do the mapping were defined, where the 'mapping' refers to mapping the values in the metadata using the 4 schemas.

1. **One-to-one mapping (1:1)**. Some values in the metadata can be mapped directly as a value to a property of an instance. For example, the *topic* of an imprint is depicted as code *650 a* in the metadata. This topic is mapped 1:1 by making it an instance of category like *newsbank:Economics_and_Trade*, and depicted here in the automatically generated RDF/XML code.

```
<newsbank:Economics_and_Trade rdf:ID="Accounting">
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Accounting</rdfs:comment>
  <swhi:subject xml:lang="en">Accounting</swhi:subject>
</newsbank:Economics_and_Trade>
```

2. **One-to-many mapping (1:m)**. Sometimes, one value can be split up into multiple properties. This is the case for the instances of *Time* and the names of a person. For example, a name written in the form `<Lastname, Firstname>` with code *100 a* can be split up by mapping the `Lastname` to *foaf:surname* and the `Firstname` to *foaf:firstName*.
3. **Many-to-one mapping (m:1)**. Sometimes, information contained in several subfields can be concatenated to one value of one property. Datafield *300*, for instance, describes the physical properties of an imprint. The values of these subfields can be concatenated and given as value for *dc:format*.
4. **Filtering redundant and non-descriptive knowledge**. This example shows the location *Boston* with *260 a*.

```
<datafield tag="260" ind1="" ind2="">
   <subfield code="a">[Boston] :</subfield>
   <subfield code="b">N. Coverly, Jr. printer,
      Boston.,</subfield>
   <subfield code="c">[between 1810 and 1814]
   </subfield>
</datafield>
```

However, that information has also been entered by a librarian in datafield *752* of the same record in a more informative way, i.e. it has a country, state, city combination. This means that code *260 a* is redundant in our case and does not need to be mapped. Besides redundant information, there is non-descriptive knowledge in the form of administrative metadata, which is not useful for historians or other non-librarian users.

**Time and Events** Since we have a historical ontology, the method to link any object with *time* is crucial. Time can be presented in a 'discrete' and 'conceptual' view. The former is expressed in the metadata with the unit *year*, and the latter is expressed as an event. For example, *Queen Anne's War* is a conceptual expression of time, standing for the linear temporal interval of `1702-1713`, which is identified as an *event* for the ontology, thus instance of *vicodi:Event*.

```
<datafield tag="651" ind1="" ind2="0">
   <subfield code="a">United States</subfield>
   <subfield code="x">History</subfield>
   <subfield code="y">Queen Anne's War, 1702-1713</subfield>
   <subfield code="v">Personal narratives.
      </subfield>
</datafield>
```

The algorithm to extract *Time* and *Events* for code *651 y* is:

**Algorithm** *processTime*(T)
(∗ Extracting *Time* for code *651 y* ∗)

```
if T contains a question mark
  then T is instance of 'vicodi:FuzzyTemporalInterval'
  else
       if T contains the pattern 'dddd-dddd' or 'between dddd and dddd'
         then T is instance of 'vicodi:TemporalInterval'
         else
              if T has pattern 'dddd'
                then T is instance of 'swhi:Year'
                else
                     if T begins with string, followed up with a number
                       then T is instance of 'vicodi:Event'
                       else  T is instance of superclass 'swhi:Time'
```

Much of our script makes use of such heuristics. Eventually, some values do not meet the condition of these rules. These are fetched with the last **else** statement as instances of superclass *swhi:Time*. This information can be further processed by making them instances of the more concrete subclasses of *swhi:Time*. Examples of the 'residue' using this algorithm:

- *in the year one thousand seven hundred and seventy-five*
- *MDCCLXXXIV*

## 3 Preliminary Results

Table 2 depicts the distribution of all the instances over the populated (used) classes of the SWHi ontology, including the populated subclasses, if applicable. It shows that 44,298 instances of *foaf:Document* were created, which describe the *Imprints*. There are actually 36,305 imprints in the metadata, but an imprint was also classified using multiple *types of imprints* (e.g. the topics 'Broadsides' AND 'Hymns') in the metadata.

**Table 2.** Number of instances for a class

| | **Instances** | |
|---|---|---|
| **Class** | # | % |
| `foaf:Document` | 44298 | 48.60 |
| `foaf:Organization` | 26634 | 29.22 |
| `foaf:Person` | 10225 | 11.22 |
| `vicodi:Time` | 7093 | 7.78 |
| `vicodi:Category` | 1909 | 2.09 |
| `vicodi:Location` | 818 | 0.90 |
| `vicodi:Event` | 163 | 0.18 |
| `vicodi:Language` | 11 | 0.01 |
| **Total** | **91151** | **100** |

The table shows that 1909 subjects are covered by the *Imprints*, and these subjects are grouped together in 16 categories. The *document collection* is classified and clustered using this topic hierarchy. There are 163 events in our historical ontology, and many more 'discrete' *time* instances. Thousands of names have been extracted (*foaf:Person*), as well as hundreds of locations (*vicodi:Location*). The former type of instance can be regarded as a short *biography* (names, dates, topics of interests, publications, etc), whereas the latter type of instance can also be seen as a very simple *gazetteer*, since it lists combinations of a country, province (or state) and capital. Besides the 91,151 instances, there are 334 direct classes, 46 direct properties and in total 1,003,180 statements. About 50 MARC21 codes and up to 46 properties are used, and it is about 100 MB big.

The quality of the populated historical ontology is also evaluated by exploring its potential to answer user queries. We stored the ontology in Sesame[9], which

[9] `http://www.openrdf.org/`

is an open source framework for storage, inferencing and querying of RDF data. It was queried for a number of conceivable questions about the Early American Period and the Evans dataset in the RDF query language SPARQL using the subject-predicate-object principle[10]. The ontology allows us to answer the question posed in the beginning of this paper as case in point: *"Did George Washington publish about the events that have only occurred in his life?"*

The results of this query are shown in Table 3. It shows that in our dataset, George Washington mostly published about the American Revolutionary War (1775-1783) during his life, which is not surprising, because he was a key actor in that event. *And who did also publish about this event besides George Washington?* The ontology returns 366 results, which were retrieved in 139 ms. A subset of the relevant nuggets of information for this question is depicted in Table 4. So we can continue providing context to the answers by traversing the RDF graphs and linking all relevant nuggets of information together with inference.

**Table 3.** Results of query: *Did George Washington publish about the events that have only occurred in his life?*

| # | **Event** |
|---|---|
| 1× | "French and Indian War 1755-1763" |
| 1× | "Washington's Expedition to the Ohio, 1st, 1753-1754" |
| 8× | "Revolution 1775-1783" |

**Table 4.** Query: *Who did also publish about the American Revolutionary War (1775-1783) besides George Washington?* The first and last results of 366 answers are depicted.

| **Name of Author** | **Lifespan** | **Event** |
|---|---|---|
| "Bancroft, Edward" | "1744-1821" | "Revolution 1775-1783" |
| "..." | "..." | "..." |
| "Mansfield, Isaac" | "1750-1826" | "Revolution 1775-1783" |

## 4 Summary and Future Work

Libraries increase their amount of metadata each day, but much of these metadata is not used as aid to disclose subjects. We have used these metadata to create a Semantic Web compatible "historical" ontology. An advantage for historians and other users is that these technologies will make implicit knowledge explicit, and new perspectives can be gained. By aligning schemas and vocabularies, historical objects are described with more semantics, additional (implicit) relationships can be explored.

[10] `http://www.w3.org/TR/rdf-sparql-query/`

For digital libraries, there are various potential benefits to adopting Semantic Web technology. A common ontology could be developed to store and maintain all metadata. That is easier said than done, but such an ontology could be developed and accepted. Adding an ontological layer to metadata makes this information much more valuable and accessible. The Protégé GUI, for instance, allows librarians to create instances and properties using forms, and may be easier to work with and less error prone than using a Wordpad-like environment, which librarians use now to compile metadata. For the development of the Semantic Web, metadata can be valuable. The metadata are carefully entered by librarians, and provides information which is hard to obtain otherwise. Existing library metadata repositories can be made useful in the (near) future in a Semantic Web context by mapping them as we have done. New ontologies can be populated using the vast amounts of already existing metadata.

The SWHi project is still in progress. We are developing semantic services based on our ontoloy, and continue to improve our ontology by populating it further with term extraction from full texts. We are also planning to evaluate our Semantic Web application with experimental empirical user studies.

## References


1. Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5):34–44, 2001.
2. Ismail Fahmi, Junte Zhang, Henk Ellermann, and Gosse Bouma. SWHi System Description: A Case Study in Information Retrieval, Inference, and Visualization in the Semantic Web. In Enrico Franconi, Michael Kifer, and Wolfgang May, editors, *ESWC*, volume 4519 of *Lecture Notes in Computer Science*, pages 769–778. Springer, 2007.
3. Deborah L. McGuinness. Ontologies Come of Age. In Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, editors, *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, pages 171–194. The MIT Press, Cambridge, Massachusetts, 2005.
4. Gábor Nagypál, Richard Deswarte, and Jan Oosthoek. Applying the Semantic Web: The VICODI Experience in Creating Visual Contextualization for History. *Literary and Linguistic Computing*, 20(3):327–394, 2005.
5. Natalya Noy. Order from chaos. *Queue*, 3(8):42–49, 2005.
6. Guus Schreiber, Alia Amin, Mark van Assem, Viktor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Zhisheng Huang, Janneke van Kersen, Marco de Niet, Borys Omelayenko, Jacco van Ossenbruggen, Ronny Siebes, Jos Taekema, Jan Wielemaker, and Bob J. Wielinga. Multimedian e-culture demonstrator. In *International Semantic Web Conference*, pages 951–958, 2006.
7. Mark van Assem, Maarten R. Menken, Guus Schreiber, Jan Wielemaker, and Bob Wielinga. A method for converting thesauri to RDF/OWL. In Sheila A. Mcilraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *Proceedings of the Third International Semantic Web Conference (ISWC 2004)*, pages 17–31, Hiroshima, Japan, November 2004. Springer.
8. B. J. Wielinga, A. Th. Schreiber, J. Wielemaker, and J. A. C. Sandberg. From thesaurus to ontology. In *K-CAP '01: Proceedings of the 1st international conference on Knowledge capture*, pages 194–201, New York, NY, USA, 2001. ACM Press.