Public collections on the Semantic web in a Hungarian context

The article offers a short introduction to the appearance of the semantic web in public collection environment. In the following, case studies are presenting various semantic web-related projects in libraries and museums from Hungary with collaboration forms to European frameworks as well. We are also discussing shortly some future project development plans in this field.

Introduction

The appearance of semantic web related to libraries museums and archives, in our view, can lead to a paradigm-shift in the fields of information search and retrieval and digital document management.

Regardless of whether information is in people's mind, in physical or digital documents, or in the form of factual data, it can be linked. Linked data is not a properly defined technical standard but an approach and set of technologies that aim to bring the benefits of the web to data, not just to documents. Linked data gives us a web of data rather than a web of documents, and it is RDF that gives linked data its basic shape (Meehan, 2014)RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (it is referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. ("Resource Description Framework (RDF)," 2014)

The presence of semantic ontologies, new type of namespaces based on RDF/XML metadata description and the several types of application that based on it (in connection with the Linked Open Data conception) are appearing as a gateway for libraries to the semantic web universe. The harmonization of the traditional data exchange standards with the new semantic web-compatible environment seems to be an essential task.

More and more libraries, museums and archives want to publish their standardized datasets on the semantic web. In order to reach this goal, they have to build up semantic ontologies. A semantic ontology is an explicit specification of a conceptualization. RDF/OWL language is appearing as a representation ways of ontologies. Metadata description and cataloguing is appearing in RDF/XML language. Other kind of standard public collection data inputs (like Dublin Core, LIDO) must be converted to that format. Namespaces can identify the different kind of data inputs that are appearing in RDF/XML environment. Thesauri and authority data can also appear on semantic web. Here we describe some major semantic web service environments.

In the SKOS environment specifications and standards are being developed to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web. Data must be published as linked open data in order to build-up standard connections with other standard RDF/XML based datasets ("Introduction to SKOS," 2012)

The Virtual International Authority File (VIAF) is an international service designed to provide convenient access to the world's major name authority files. VIAF appears as a building block for the Semantic Web to enable switching of the displayed form of names for persons to the preferred language and script of the Web user. VIAF began as a joint project with the Library of Congress (LC), the Deutsche Nationalbibliothek (DNB), the Bibliothèque nationale de France (BNF) and OCLC. It has, over the past decade, become a cooperative effort involving an expanding number of other national libraries and other agencies. (OCLC, 2016)

FOAF is a project devoted to linking people and information using the Web. FOAF integrates three kinds of network: social networks of human collaboration, friendship and association; representational networks that describe a simplified view of a cartoon universe in factual terms, and information networks that use Web-based linking to share independently published descriptions of this inter-connected world. FOAF, like the Web itself, is a linked information system. It is built using decentralised Semantic Web technology, and has been designed to allow for integration of data across a variety of applications, Web sites and services, and software systems. FOAF was designed to be used alongside other such dictionaries ("schemas" or "ontologies"), and to be usable with the wide variety of generic tools and services that have been created for the Semantic Web. The initial focus of FOAF has been on the description of people, since people are the things that link together most of the other kinds of things we describe in the Web: they make documents, attend meetings, are depicted in photos, and so on. The FOAF Vocabulary definitions presented here are written using a computer language (RDF/OWL) that makes it easy for software to process some basic facts about the terms in the FOAF vocabulary, and consequently about the things described in FOAF documents. A FOAF document, unlike a traditional Web page, can be combined with other FOAF documents to create a unified database of information. FOAF is a Linked Data system, in that it based around the idea of linking together a Web of decentralised descriptions. (Brickley & Miller, 2014)

Following the replacement of AACR2 standard by RDA, BIBFRAME is widely viewed as the replacement for MARC as a data exchange standard framework. Much like MARC, the Library of Congress initiated it. BIBFRAME is an abbreviation – not an acronym despite the capitalisation – for the BIBliographic FRAMEwork Initiative. (Meehan, 2014). The new bibliographic framework project is focusing on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model. ("A Bibliographic Framework for the Digital Age," 2011). The implementation of BIBFRAME to library environment has just started, the first results expected to appear soon.

1. The Hungarian National Library: First national semantic web project

The National Széchényi Library (NSZL) published its entire OPAC and Digital Library and the corresponding authority data as Linked Open Data in 2010 as one of the first public collections in Europe. The used vocabularies are RDFDC (Dublin Core) – (MARCX ML to RDF/XML conversion with XSLT for OPAC bibliographic data) FOAF for names, and SKOS for subject terms and geographical names. NSZL uses CoolURIs. Every resource has both RDF and HTML representations. The RDFDC, FOAF and SKOS statements are linked together. The name authority dataset is matched with the DBPedia (semantic version of Wikipedia) name files. NSZL also supports the HTML link auto-discovery service. All of the available linked data resources (names, subject authority, catalogue records) can be searched and retrieved to external resources in the semantic web via an SPARQL endpoint with specific browser tools (that are useful in machine-to machine communication). (Horváth, 2011b)

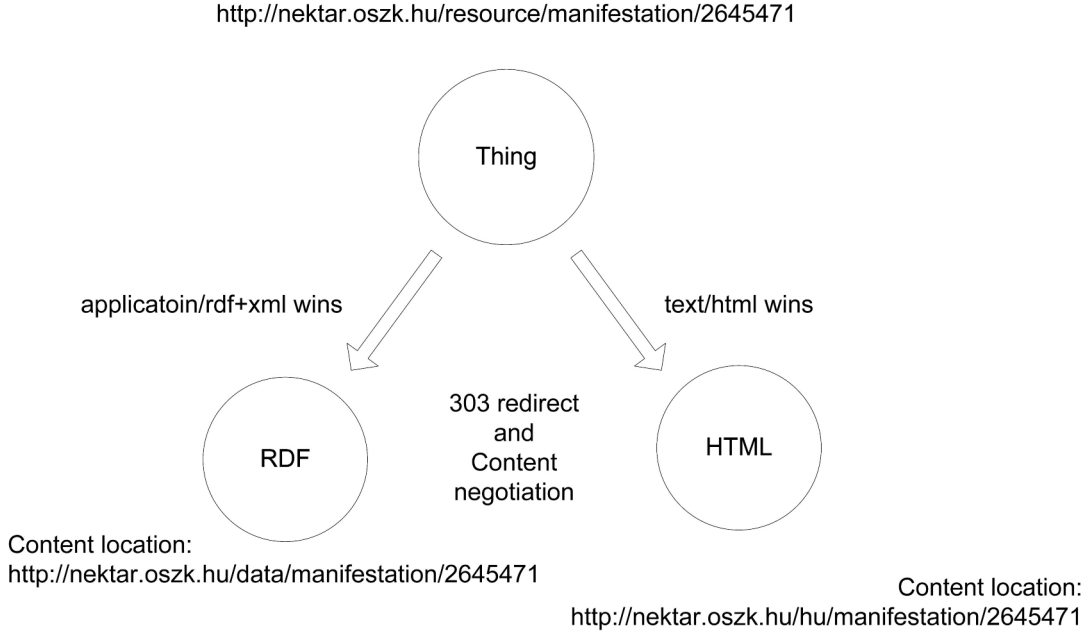
There was no specific project related to this field in the library. Small developments pointed to the same direction. Three members of the directorate of informatics developed it when time permitted. In 2009 they realized that they almost had everything in order to publish linked data on the semantic web. They converted the library thesaurus to SKOS format. Via the LibriURl tools the OPAC records have become accessible via URL. The URL-based search in the NSZL integrated system have become available via the SRU protocol with the Yaz proxy tool. They could use the experiences of the Swedish National Library (LIBRIS) semantic web implementation project. The main aims were the following: Library datasets need to be open (get your data out,) need to be linkable, and also need to provide links. Datasets must be part of the network, cannot be an end in itself and the system must allow hackability.

The major advantages of the RDF-based semantic model are the following:

- RDF Clients can look up every URI in the RDF graph over the Web to retrieve additional information
- Information from different sources merge Information from different sources merge naturally
- RDF links between data from different sources can be set
- Information expressed in different schema, can be represented in a single model

(Horváth, 2011a)

The model can be described simply in the following way. You can find the content location and the representation ways of the manifestation of the content in RDF and HTML formats:



(Horváth, 2011a)

– If application/rdf+xml is accepted the xml is given from this address via content negotiation and 303 redirect: http://nektar.oszk.hu/data/manifestation/2645471

– If text/html is accepted: Depending on the language of the browser either the Hungarian or the English interface of the OPAC (LibriVision) is given.The default is Hungarian (again via content negotiation):
http://nektar.oszk.hu/hu/manifestation/2645471

```
– <rdf:RDF>
  – <foaf:Person rdf:about="http://nektar.oszk.hu/resource/auth/33589">
      <dbpedia:birthYear>1825</dbpedia:birthYear>
      <foaf:name>Mór Jókai</foaf:name>
      <foaf:name>Jókai Mór</foaf:name>
      <foaf:name>Jókai Mór (1825-1904)</foaf:name>
      <dbpedia:deathYear>1904</dbpedia:deathYear>
      <owl:sameAs rdf:resource="http://dbpedia.org/resource/M%C3%B3r_J%C3%B3kai"/>
      <foaf:givenName>Mór</foaf:givenName>
      <foaf:familyName>Jókai</foaf:familyName>
    </foaf:Person>
</rdf:RDF>
```

(Horváth, 2010)

Different URI -s can be used for the same resource. For example Mór Jókai Hungarian writer (born in Komárno) can be identified with two URI-s http://nektar.oszk.hu/resource/auth/33589 in the NSZL library database and http:// http://dbpedia.org/resource/M dbpedia.org/resource/M in DBPedia. The owl: SameAs links are resolving this problem. With this kind of link DBPedia can attach the VIAF link of the same person to its semantic interface and the different language versions of the same entry also (see at: *http://dbpedia.org/page/M%C3%B3r_J%C3%B3kai*).(Horváth, 2010)

The collection of Hungarian National Library in this way have become a part of the semantic universe. The next step was to help other public collections to convert data to semantic web format and publish them also in the semantic web. That was the exact goal of the EU financed ALIADA project that will be described shortly in the following.

2. ALIADA Project – a short overview

The international ALIADA project, financed by the EU focused on creating tools for the implementation of semantic web to public collection environment. The project partners came from different European countries and different sectors (software developing industry, library, museums, and archives). (Horváth, 2015) The starting point is easy in a sense that GLAM (Galleries, Museums, Archives and Libraries) institutions have usually rich metadata related to their collections. Metadata mainly stored in standard schemes (MARC, Dublin Core, Lido etc.) Through the ALIADA framework, various metadata subsets from library, archives, gallery museum management systems (GLAM catalogue data, bibliographic data and authority data) can be converted from standard metadata input forms (e. g. MARC, LIDO, DUBLIN CORE) into RDF based semantic compatible format according to the ALIADA ontology. (Aliada Project, 2015) The conversion process is being made with an open source Java software. Data subsets are being stored in a Virtuoso database and exported to a datadump file that is publicly available online. All the semantic data subsets through a SPARQL endpoint are being registered in the datahub.io database with standard descriptions, links to the subsets and the address of the semantic Virtuoso database. Even before the automatic publication of semantic datasets in the semantic cloud, these can be linked to other datasets. The ALIADA software also automatizes the whole conversion and publication process. The partner institutions have to provide only standard metadata input subsets; the public collection experts do not need deep expertise on semantic web technologies. The semantic datasets can be linked to other datasets, such as Europeana, British National Bibliography, Spanish National Library, Freebase Visual Art, DBpedia, Hungarian National Library, Library of Congress Subject Headings, Lobid, MARC codes list, VIAF Virtual International Authority File or Open Library. (Ádám Horváth, 2014)

The project just finished in October 2015. The ALIADA software tool is free and publicly available to all the interested parties under the terms of the GNU GPL v3 (Ádám Horváth, 2014; Aliada Project, 2015).

An example of practical use of ALIADA framework tool will be described in the next chapter.

3. A case study of a semantic web related project in the Petőfi Literary Museum with an integrated library system in museum environment

An example of practical use of a semantic web based database is to build a triple store from a part of the database of the Qulto integrated library and museum automation system of the Petőfi Literary Museum (PIM, the abbreviation of its Hungarian name: Petőfi Irodalmi Múzeum). The museum's duty is to collect documents and objects connecting to the important personalities of the Hungarian literature. The museum's library also collects the documents of this theme, and the bibliographic descriptions of the library catalogue use the records, and the descriptive metadata of the museum inventory items. These catalogue items also contain important additional information about the novelists who are, as authors or mentioned personalities are joined to the museum records. (Bánki & Mészáros, 2016)

The common information contained by a name authority record in a library catalogue or in a museum electronic inventory system, are personal name, date of birth and death, title, profession, data sources and linked bibliographic data, but there are a plenty of attributes in the catalogue of the library automation system of PIM, added to the records e. g. prices, exact date of birth and death, place of birth, death and living, parents, husband, wife, children, sex, religion, education, workplaces, important events of the life of the novelist etc. From these attributes a complex information packet has prepared and stored in the Qulto integrated collection management system of the PIM. Most of these information had been imported not in the Qulto ICMS, but in 22 separate Access databases, which were used by the experts of the museum before Qulto system have been introduced for ten years. A data conversion was necessary from the Access based system to Qulto, and after the migration the information had been added to separate authority records. These information units had to be merged into one, main record from the various data items. In three steps more than 110 000 name authority record were selected as duplicated record. Duplication means that another name authority record was found in the database of PIM as a main name record describing the same person. After the information could be merged from the duplicated record to its main pair, the corrected database was ready to become one of the base data store elements of the Hungarian National Namespace. At the same time, it has also published on semantic web. So after consolidating the name database, and the record number was decreased to 620 000 items in the Qulto database of Petőfi Literary Museum, the dataset was ready to be uploaded to somewhere or to be prepared as a local triple store.

There were three possibilities for us, to publish the authority records of the collection Management System of Petőfi Literary museum on semantic web.

1: Load it to VIAF

2: Build a triple store working together with other Hungarian museums, for example, Hungarian National Museum, or Museum of Fine Arts.

3: Create an own triple store in the PIM.

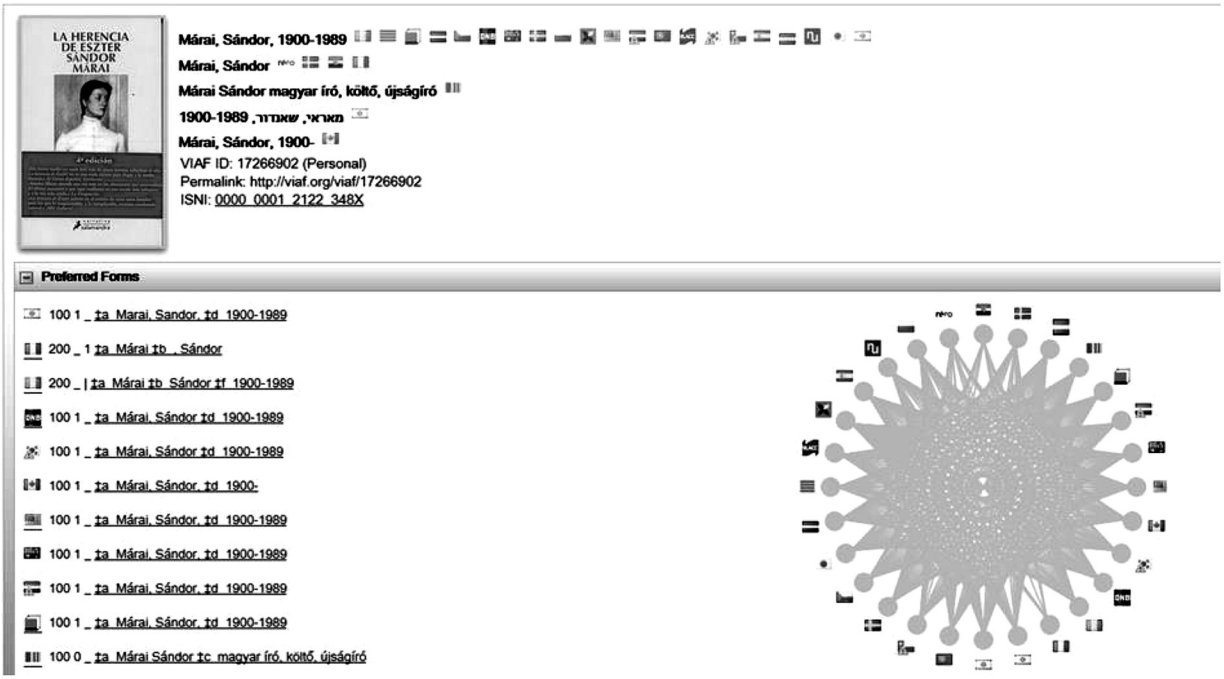
We describes each option in the following sub-chapters.

3.1: Load to VIAF

An option for publication is to connect to OCLC and load it to the VIAF database. As we have already mentioned above The VIAF – Virtual International Authority File – is, as an important unit of semantic web, coordinated by OCLC. It has based mostly on the authority records of libraries, so it works like a library catalogue in this sense. The identification of personalities is based mostly not on the metadata of the name records, the personal and biographic information of the novelists, but on the bibliographic data linked to them, the documents were written by or about them. This way of identification is convenient for a library, having usually the name records of authors, but not for the factographic database of a museum. These institutions have not much books, but have information entered from reference books. They have to identify the units of the authority database, the persons themselves, by the attributes of the biographic data.

Otherwise the upload is useful, necessary, and hopefully VIAF can use the uploaded records. By first step we've connected to OCLC and sent a dataset of a first trial version of authority data export file, containing authority records having already linked bibliographic data in the local library catalogue. The VIAF needed a MARC21 export, that should have been prepared, creating a HUNMARC – MARC21 conversion from the Qulto ICMS, which as a MARC based system uses the Hungarian national standard HUNMARC as internal data storage format.

The VIAF has already got a plenty of name data form Hungary, hopefully these name elements will be automatically identified by the system of VIAF, and the already in the VIAF database existing records will be enriched by the newly sent data, and also new records will be created from the personal database sent from PIM to VIAF by this step. Therefore, creating an authority export for VIAF upload, the personal names were selected, which had bibliographic records in the database, and had enough additional information as authority records too.



3.2: Using the ALIADA application that has already installed in Hungary

We have already written a short overview about the ALIADA application (software tool framework) in the previous chapter. Here we are offering a practical example of its use.

The Museum of Fine Arts of Budapest has built its own Aliada database, with the possibility to define more sub users and sub databases. (http://www.szepmuveszeti.hu/aliada_en). The museum has published the descriptions of its 4000 artefacts on the semantic web with the help of ALIADA tool, and also gave the possibility to the Petőfi Literary museum, to try this application, both the input and the web based public interface.

The workflow of data upload was by the ALIADA pilot project the same as by the OCLC. First we had to choose the records to be uploaded. The aspects of selection were almost the same, so the records had to contain enough information, they had to be entered into the proper sub databases. It is possible to upload to ALIADA those authority record, which have not any bibliographic records joined to them. Thanks to the six years long joint work of the experts of the museum and library and the Qulto software support, the redundancy of the database has been almost fully decreased. On the other hand, it was necessary to control and filter the duplicates of the uploaded names from the database in ALIADA. The existence of obligatory data elements had to be checked also. As in the case of VIAF export some data manipulation was necessary, e.g. the bibliographic data links were filled also into the authority data, to make it the integral part of the MARC authority record.

The Qulto internal format based on the structure of MARC, and it has its own structural logic, so the authority data have their quasi authority elements. For example, an authority record can be joined to corporate or geographical name records in a hypertext seeming data network. All these attached sub authority elements had to be appended to the authority output, and

a proper MARC 21 header should have been prepared by the MARC authority export as well. In the past 6 years the PIM personal authority records were developed to be able to contain a plenty of various information, in various MARC fields and subfields, not defined in the default MARC21 standard. These new data elements had to be mapped to the MARC21 data fields, being recognizable for ALIADA MARC21 import format. In the future we'll try to enhance the acceptable field list of ALIADA MARC import. During the MARC import ALIADA converts the authority MARC data to RDF statements. ALIADA is a user- friendly and easy to use application. The operator has to validate the input data set, has to select the sub database (graph), and delete the unnecessary records from the Virtuoso database. The ALIADA import program always adds elements, but never merge duplicated records. You have to filter your dataset from duplicated records before the ALIADA import! If necessary, you can select the demanded data type, and mark the data fields and subfields to be converted through the import process. There is a problem by import in the pilot project: a relatively small size of input files have accepted by ALIADA.

The result is the converted dataset, into the Virtuoso database, which is browseable, containing valid data links generated by ALIADA. The dataset can be insert to the semantic cloud. Another possibility is, to join data elements automatically with other ones, and these links can be added to the local database, to enrich authority or bibliographic records with other data connections. Also the VIAF URI-s can be added to the authority records in the local database.

Our goal is to further enhance this semi-automatized workflow (that has built from these four steps: 1. data manipulation in PIM Qulto database 2. data conversion from Hunmarc to MARC21 3. preparing MARC XML from relational database quasi MARC data units 4. Aliada import converting to RDF statements of Virtuoso database) to develop a fully automatized one. Also the VIAF upload is planned to be automatized.

3.3 An own triple-store in PIM

The third possibility, to build an own triple store of PIM, potentially means to install an own Aliada application to the local server of Petőfi Literary Museum. After the RDF statements were controlled, the new database is ready several output formats to be prepared from it: FRBRoo, WGS84, SKOS, SKOSXL, FOAF, DCTERM, OWL-TIME.

The advantages of Qulto as a Library or Museum Collection management software, is the almost unlimited possibility of defining new special data elements, and also the highly customized segmentation of the records. So all the various information segments can be added in separate and specially marked record fields. In this way any type of outputs, and export data sets can be produced from this input of records.

The potential aim of use of semantic web databases and database elements, is to identify and describe persons who are hardly definable by name strings, or have many connections and are enriched with plenty of sub elements.

An example is below for the famous Habsburg emperor Joseph the Second, having a well-known and often mentioned but very short name which is hard to identify by entering search terms, and has a long name with plenty of titles and Christian names on the other side. The record is (for he was emperor of the Holy Roman Empire, so the head of Germany that time) for Deutsche Bibliothek tries to identify him.

Link zu diesem Datensatz http://d-nb.info/gnd/118558404

Person Joseph II., Heiliges Römisches Reich, Kaiser

Geschlecht männlich
Andere Namen Joseph II., Römischer Kaiser
Joseph II., Deutschland, Kaiser
Josef, Österreich, Erzherzog, 1741-1790
Josef II., Heiliges Römisches Reich, Kaiser
Joseph II., Heiliges Römisches Reich, König
Joseph, von Habsburg-Lothringen
Josephus II., Heiliges Römisches Reich, Kaiser
Josephus II., Imperium Romanum-Germanicum, Imperator
Giuseppe II., Imperio Romano-Germano, Re
Augusto Guiseppe II., Imperio Romano-Germano, Re
Joseph II., der Grosse
Joseph, der Zweite
Joseph, der II.
Joseph, II.
Josephus II., Imperator
Giuseppe, d'Austria
Joseph Benedikt, Prinz
Joseph Benedikt August Johann Anton Michael Adam, Österreich, Erzherzog, 1741-1790
József II.
Josip II.
Graf Falkenstein (Pseudonym)
Falkenstein, ..., Graf (Pseudonym)

Quelle Internet (Stand: 07.08.2014): https://de.wikipedia.org/wiki/Joseph_II.
LoC Auth
DbA (WBIS)
M; B 1986
Zeit Lebensdaten: 1741-1790
Wirkungsdaten: 1765-1790
Land Österreich (XA-AT)
Geografischer Bezug Geburtsort: Wien
Sterbeort: Wien
Beruf(e) Kaiser
Funktion(en) Herrscher
Weitere Angaben 1765-1790 Kaiser (bis 1780 als Mitregent Maria Theresias)
Beziehungen zu Personen Isabella, Österreich, Erzherzogin (erste Ehefrau)
Maria Josepha, Heiliges Römisches Reich, Kaiserin (zweite Ehefrau)
Maria Theresia, Österreich, Erzherzogin (Mutter)
Franz I., Heiliges Römisches Reich, Kaiser (Vater)
Maria Theresia, Österreich, Erzherzogin, 1762-1770 (Tochter)
Systematik 16.5p Personen der Geschichte (Politiker und historische Persönlichkeiten)
Typ Person (piz)

4. Other semantic-web based project plans from the Hungarian library sphere

As the software distributor of Qulto Integrated Collection Management System, Monguz Ltd has a leading role to help building national aggregation and shared cataloguing systems for its customers. All the databases of the important national and

international projects listed below are suitable to function as authorized data source for semantic web in order to fulfil the demands of the end users to get controlled and relevant data from the national databases via their own content management system.

The main projects, where semantic web development work can be in progress: MOKKA, the National Hungarian Shared cataloguing system; Museummap, the national aggregator system of Hungarian museums for the Europeana project; ELDORADO, which's main objective is to provide digital contents from Hungarian libraries cooperating with publishers, with respect to copyright issues; ODR, the Hungarian National Document Supply System; and a Polish project that can be relevant example in order to create semantic datasets through the Hungarian museum aggregation project: the Museum Portal of NMK, National Museum of Krakow, a common search interface for the museums of the region of Small Poland (Malopolska), led by the National Museum of Krakow.

5. Schema.org and microdata: New semantic web tools in the HTML5 standard

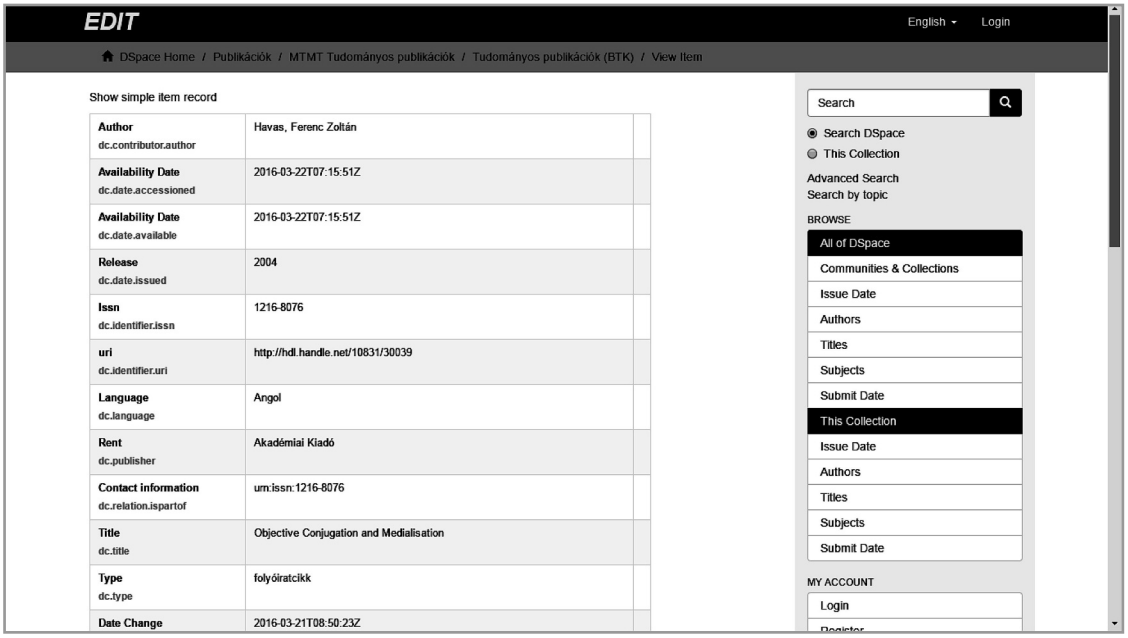
Many librarians are familiar with basics of the HTML language. Usually, HTML tags tell the browser how to display the information included in the tag. For example, <h1>Avatar</h1> tells the browser to display the text string "Avatar" in a heading 1 format. However, the HTML tag doesn't give any information about what that text string means – "Avatar" could refer to the hugely successful 3D movie, or it could refer to a type of profile picture – and this can make it more difficult for search engines to intelligently display relevant content to a user. The web of documents is linking documents links are not qualified. Otherwise on the semantic web we are linking datasets with qualified links. Schema.org simply provides a collection of shared vocabularies that can be used to mark up the public collection homepages (and any other homepages of course) in ways that can be understood by the major search engines: Google, Microsoft, Yandex and Yahoo! You can use the schema.org vocabulary along with the Microdata, RDFa, or JSON-LD formats to add information to your Web content.("Getting started with schema.org using Microdata," 2016) In case of RDF-a, the RDF statements are properties of HTML tags and can be generated as a collection of HTML-based homepage texts.

Why microdata and microformats are useful? The web pages have an underlying meaning that people understand when they read them. But search engines have a limited understanding of what is being discussed on those pages. By adding additional semantic tags (for example with RDFa format) to the HTML of your web pages – tags that say, "Hey search engine, this information describes this specific movie, or place, or person, or video" – you can help search engines and other applications better understand your content and display it in a useful, relevant way. Microdata is a set of tags, introduced with HTML5, that allows you to do this. (Horváth, 2016)

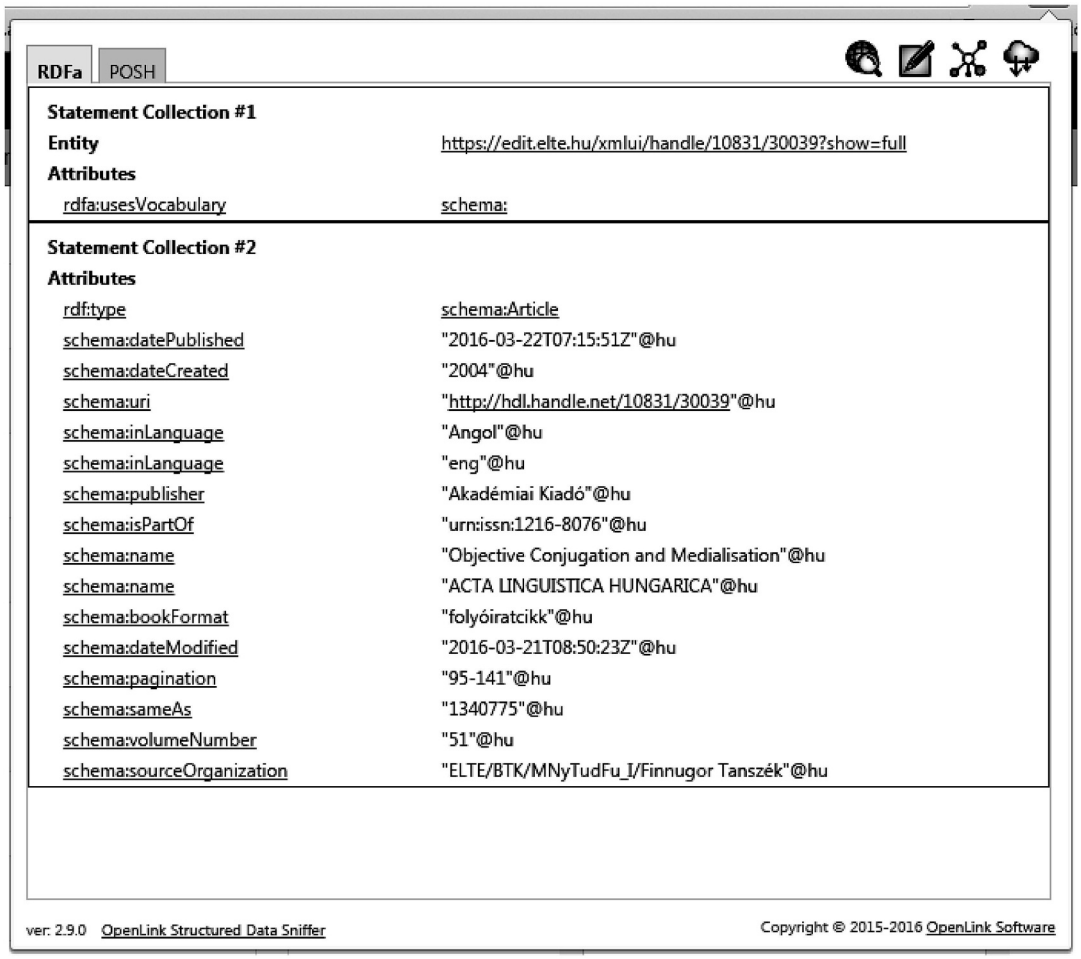
In Libraries with the help of Schema.org you can use the Library class and define FRBR-like attributes on the homepages (exampleOfWork, workExample). It is possible to define also connections (hasPart, isPartOf). Currently microformats (schema.org and RDF-a) are being used in OPAC (WorldCat, Koha), and in discovery systems (like VuFind), and repositories (like DSpace).

In Hungary the first implementation of microformat tags can be found in the university library of the most traditional university in Budapest, Eötvös Loránd University (ELTE). The pages of the Dspace-based institutional repository: ELTE Digital Institutional Repository (EDIT) have tagged with RDFa and Schema.org tags. Microformats will be used soon also in the Vufind based new integrated portal of the Hungarian National Library (support of microformats is a built-in function of VuFind). (Horváth, 2016)

Here are some examples:



A sample record with semantic microformat tags in EDIT repository



(Horváth, 2016) Sample of semantic statements in shema.org and RDF-a

6. Scientific manuscripts on the semantic web

A rather new development is the appearance of semantic data producing and aggregation in the field of digital philology. The methods and tools can be basically the same as we describe above regarding to museum and library case studies.

The Petőfi Literary Museum is a partner in the Digitised Manuscripts to Europeana project ("DM2E: Digitised Manuscripts to Europeana," 2016). The project has five work packages. The first is focusing on the content integration in a semantic way to Europeana, the second one will provide the interoperability infrastructure for translating content from its current source formats into the Europeana Data Model (EDM) based on already existing tools. The third working package has a focus on validation of the results. The fourth one based on the dissemination and community building and fifth one on the project coordination.

New recommendations will be provided to digitisation methods of manuscript and the method to enrich them with metadata as well. The second step will be the publication of metadata on the semantic web together with metadata conversion and aggregation. The DM2E model will support document hierarchies, the use of different ontologies. It will support a proper Uri syntax and the enrichment of the local manuscript databases from semantic resources. (Fellegi, 2016)

The metadata aggregation in the Europeana Data Model has based on a proper RDF namespace, semantic RDF files can be generated via XSLT transformation. Data exchange via the different public collection targets can be managed via OAI-PMH protocol. Europeana provides an internal metdata quality check protocol in order to make full compliance of the different imported datasets with the Europeana data model. Data enrichment options are also available via VIAF and DBPedia.

The DigiPhil (Online Knowledge Base of Scholarly Text Editions, Bibliographies and Researcher Repositories) project (DigiPhil project, 2016) of the Petőfi Literary Museum can help to compile a list of manuscript metadata in collaboration with several research groups from Hungary and to transform the transcripts of texts into mark-up language compliant text forms (that can be published on the web even enriched together with semantic microdata in the future). Long-term preservation, platform-independent search functions, and aggregation can be ensured by using standards, both in transcription and metadata description. The DigiPhil project uses internationally accepted standards for these reasons. The mark-up language transcription follows the Text Encoding Initiative guidelines, bibliographic data is given according to the MARC21 standard, while the structure and the syntax of the metadata follow the workflow developed by Digital Manuscripts to Europeana (DM2E).

Conclusion

We can summarize all the efforts the public collections are making in the semantic web field that the main aim is to provide even more effective online visibility and enrichment of our datasets and contents. We can share our datasets on the sematic web, be a part of the semantic cloud, enrich our collection from external resources. On the other hand the use of persistent URL-s , and the RDF-a/Schema.org microformats in HTML 5 environment can help to retrieve data from our web based content in a much more comprehensive way then it was possible before.

Special thanks to our colleagues for the professional contribution to this article: Ádám Horváth (Central Library, Hungarian National Museum). Gábor Simon, Ádám Pogány (Hungarian Museum of Fine Arts), Zsófia, Fellegi, Anikó, Mohay, Zsolt Bánki, Gábor Palkó (Petőfi Literary Museum).

Bibliography

A Bibliographic Framework for the Digital Age. (2011). Retrieved May 3, 2016, from http://www.loc.gov/bibframe/news/framework-103111.html

Ádám Horváth. (2014). The European ALIADA project (pp. 1–20). Rome: 33rd ADLUG Annual Meeting. Retrieved May 3, 2016 from http://www.slideshare.net/aliadaproject/aliada-intro-adamhorvath03

Aliada Project. (2015). Introduction to ALIADA webinar. Retrieved May 3, 2016, from http://www.slideshare.net/aliadaproject/introduction-to-aliada-webinar

Zsolt Bánki, & Tibor Mészáros (2016). *Checking the identity of entities by machine algorythms is the next step to the National Name Authorities.* Retrieved May 3, 2016 from https://conference.niif.hu/event/5/session/14/contribution/26

Dan Brickley, & Libby Miller. (2014). FOAF Vocabulary Specification 0.99 Namespace Document 14 January 2014 – Paddington Edition. Retrieved May 3, 2016 from http://xmlns.com/foaf/spec/

DigiPhil project. (2016). Retrieved May 3, 2016, from http://digiphil.hu

DM2E: Digitised Manuscripts to Europeana. (2016). Retrieved May 3, 2016, from http://dm2e.eu

Zsőfia, Fellegi. (2016). *Metadata description of scholarly text editions – data enrichment, aggregation, transformation*. Retrieved May 3 2016 from https://conference.niif.hu/event/5/session/10/contribution/54

Getting started with schema.org using Microdata. (2016). Retrieved May 3, 2016 from http://schema.org/docs/gs.html

Ádám, Horváth. (2010). *Linked Data at the National Széchényi Library:road to the publication*. Retrieved May 3 2016 from http://swib.org/swib10/vortraege/swib10_horvath.ppt

Ádám Horváth. (2011a). Linked Data at NSZL. Retrieved May 3 2016 from http://nektar.oszk.hu/w/images/0/04/LinkedDataAtNszl_06.pdf

Ádám Horváth. (2011b). National Széchényi Library Semantic Web wiki. Retrieved May 3, 2016, from http://nektar.oszk.hu/wiki/Semantic_web

Ádám Horváth. (2015). ALIADA as an Open Source solution to Easily Published Linked Data for Libraries and Museums. Retrieved May 3 2016 from http://www.slideshare.net/aliadaproject/swib15-aliada

Ádám Horváth (2016). *RDFa – schema.org: unity of document and semantic web*. Retrieved May 3 2016 from https://conference.niif.hu/event/5/session/10/contribution/27/material/slides/0.ppt

Introduction to SKOS. (2012). Retrieved May 3, 2016, from https://www.w3.org/2004/02/skos/intro

Thomas Meehan. (2014). The impact of Bibframe. *Catalogue & Index*, (177), 2–16. Retrieved May 3 2016 from http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=110055753&site=ehost-live

OCLC. (2016). VIAF. Resource Description Framework (RDF). (2014). Retrieved May 3, 2016, from http://www.w3.org/RDF

Márton Németh
nemethm@gmail.com ■
(Phd Student, Doctoral School of Informatics, University of Debrecen, Hungary, Public Collection Expert, Monguz Ltd, Budapest, Hungary)

András Simon
(ICMS consultant, Budapest Monguz Ltd, Hungary)