Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA

Federal Office of Culture FOC
Swiss National Library NL

Patrice Landry, 31 July 2014

MACS report on the state of MACS linked data

Context
This report was prepared as a background paper to a meeting to be held during the IFLA 2014 WLIC in Lyon. It present the current situation of the MACS project.

MACS in TEL (The European Library)

In December 2011, the TEL Management Committee approved the recommendations from the CENL Working Group on the Integration of MACS into The European Library that "the MACS data, its future technical development and expansion be integrated in The European Library". The Swiss National Library, as the MACS Project leader, continues to ensure its support to implement the TEL Management Committee's decision to fully integrate MACS data and activities in the TEL Office. In 2013, TEL Office supported maintenance and development work on the MACS links database. In particular, work in renewing the hosting of the database was done by TEL Office. TEL has also worked in converting RAMEAU identifiers used in MACS to the ARK identifiers currently used by the BnF. A sample of the MACS records augmented with the ARK identifiers was sent to DNB.
Each mapping in MACS was linked to the LOD datasets of the Library of Congress (http://id.loc.gov/authorities/subjects), BnF (http://data.bnf.fr/) and DNB (http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html)
See TEL Summary report – Appendix 1

Issues
- While TEL Office has been supportive of the MACS project, reporting has been limited and concrete results (in terms of applications) are still not visible or known.

MACS links database (LMI)
The links are produced, maintained and updated in the MACS links database (LMI) currently hosted in the Hostice server located in the Netherlands and supported by the TEL Office. An initial study of the migration of the database to the TEL's ULCC server has indicated that the LMI database could not easily be migrated to the TEL's ULCC server. This situation was further clarified in May 2013 when the TEL Office acknowledged that the LMI needs extensive root rights on the servers that prevent the move to ULCC (TEL official server). Since the MACS database is not a formal "commercial" production system, there would be difficulties in using it as a production system in the ULCC server. The decision was thus made to maintain the current hosting at Hostice as the best solution in the immediate future.

In 2013, a maintenance and software development contract was submitted to Jeroen Hoppenbrouwers, the developer of the LMI, to be financed jointly by NL and DNB. There had not been any maintenance contract since 2010 and there was an urgent need to ensured that monthly updates of headings from the BnF, DNB and LoC were processed correctly. In 2010, the auto-updating system did not process the updates correctly which corrupted some links (which were manually restored).

Maintenance work was undertaken in April 2013 and will conclude in August 2014. Delays in the work plan were due to unforeseen work commitments by the contractor and difficulties in configuring the loading of the newly redesigned GND headings. As of July 31, most of the contract tasks have been completed with success. The new LMI release includes a new efficient updates function to control and accept incoming weekly / monthly updates of headings from the BnF, DNB and LoC.

Issues
- The LMI was initially developed in 2000 using the PHP language. The PHP language is not used in TEL development work which prevents any work by TEL Office on the development and maintenance of the LMI. The current

developer of the LMI currently works and lives in the United States which makes it difficult to manage the work on a regular basis.

- In light of semantic web technology and other technical developments, new scenarios for multilingual access should be considered to improve users' access to information. As well, automatic linking procedures and software are now available that would allow to add new languages and improve current MACS languages links creation.
- The sustainability of the MACS data is a critical issue. The current MACS Links database (LMI) is 14 years old and while it is a reliable database for the storing and management of links, does not offers the semantic web and automatic linking features necessary for further development of linked data

State of linked data

The MACS project has achieved its goal in links creation by providing a critical mass of links in three subject headings languages: LCSH, RAMEAU and SWD. At the end of July 2014, there were 67'594 links with SWD, 100'924 links with LCSH headings and 126'107 links with RAMEAU headings.

The creation and maintenance of links is the responsibility of each partner. The bulk of the RAMEAU – LCSH links were provided by the Bibliothèque nationale de France in 2008 when the MACS production database was validated. Since then, the RAMEAU headings have been updated through its monthly updates until 2010. The LCSH headings in the links are updated manually when the links are updated, usually when a German GND heading is added to the link.

Most of the links created and updated were done by the Swiss National Library and the Deutsche Nationalbibliothek. Most of the time, the GND is added to an already created RAMEAU-LCSH link and the RAMEAU and LCSH are modified if out of date. The integrity of RAMEAU and LCSH headings is thus preserved. When a new link is created directly from a GND heading, the appropriate RAMEAU and LCSH headings are selected from the list of headings. MACS procedures state that when a combination of terms (main heading and subdivision) is made in RAMEAU or LCSH, adherence to the indexing rules must be respected so that searching will be made possible. Indexers are encouraged to check the BnF and LoC databases and authority files to check the validity of the constructed headings.

Issues
- The number of new links has not progressed significantly since 2011.

Use of MACS linked data

An older selection of MACS data (from before 2010) is used for multilingual searches in the Advanced Search function in The European Library portal (http://www.theeuropeanlibrary.org/tel4/) – as a prototype. It appears that the search functions have not been improved since that time.

At the end of 2013, Die Deutsche Nationalbibliothek had integrated 1 to 1 MACS linked data containing a GND heading in the GND authorities (formerly SWD authorities).

The Swiss National Library ensure a minimum multilingual search in its online catalogue Helveticat (www.helveticat.ch) by using the MACS data integrated in the GND headings. Work is underway in developing a new OPAC which will improve the multilingual search functions.

The Swiss meta catalogue, Swissbib (https://www.swissbib.ch/?lng=en) has also integrated the GND

heading in their database and currently provide access to GND subject headings by using Rameau or LCSH search terms. It is hoped that this initial development will be further developed.

In 2013, the MACS project collaborated with the "1914-1918 Online. International Encyclopedia of the First World War" (http://www.1914-1918-online.net/) by providing links in the areas of study. The MACS data are used as multilingual equivalence information as support for the librarians who index the resources.

Existing alignments between RAMEAU, LCSH and SWD (GND) in the semantic web (BnF will provide information at the meeting)

MACS linked data was used in the TELplus project

Requests have been received from the "Free your metadata Research Project" and from the Haute École de Genève's Rodin project. These requests have not been fulfilled as MACS data have not yet been published as LOD.

There has also been interest from Romania and Belgium (ANET Library Networks)

Issues

- With the new updates control now in place and the regular updating of MACS data once more possible, the elements are in place to allow the publication of MACS data as Linked Open Data. The first step will be to establish a (Creative Commons Attribution) licence that corresponds to the MACS partners' requirements.
- MACS data are used in several databases (German authority file, NL's Helveticat, Swissbib, TEL) and the question is how to ensure that MACS data are regularly updated in these databases
- The GND only takes 1 to 1 links (41'127 so far). This represents 61% of links with an GND heading
- Who else has plans?

Conclusion
The MACS project has existed and progressed over the years through the good will of its partners. They have supported the project in many ways: by their financial support for the maintenance and development of the links database, by the loading of subject headings in the links database, and by the resources they allocated for the creation and maintenance of links and for the project management.

As was pointed it out in the report, there are many issues that need to be addressed, some are technical but foremost is the question of the future of MACS. It should no longer be a project but a service. Are the partners still willing to support a service financially and to contribute to the linking work? If so, what would be the incentives to do so? If not, what to do with the data?

Appendix 1
TEL's MACS Summary activity report 2013-2014 (draft)

This year we have done:

- The RAMEU identifiers used in MACS were converted to the ARK identifiers currently used by the BnF.
- A sample of the MACS records augmented with the ARK identifiers was sent to DNB.
- Each mapping in MACS was linked to the LOD datasets of the Library of Congress (http://id.loc.gov/authorities/subjects), BnF (http://data.bnf.fr/) and DNB (http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html)
- A prototype was developed that uses MACS to link TEL's LOD records to the LOD datasets mentioned in the previous point. This prototype, therefore enriches the records with subject data in the three languages.
- The prototype was applied in the RLUK LOD dataset, enriching the dataset with the following number of links:

  Library of Congress Subject Headings - 2.388.129
  Rameau - 746.067
  Gemeinsame Normdatei - 424388

- The prototype is currently being integrated in the TEL ingestion systems for application in the TEL Open Dataset. This task is nearly finished, and will start being a part of the automatic ingestion processing of TEL during this Summer.

In the future (for the coming 12-18 months), we'd like to:

- Enrich MACS with links to Dewey Decimal Classification and Library of Congress Classification by harvesting existing links in the LOD cloud.
- Enrich MACS by integrating the multilingual subject headings that we are receiving from the libraries in project Europeana 1914-18 Collections project (includes additional languages to English, French and German).
- Represent the enriched MACS with semantic technology (RDF)
- Share the semantic representation of MACS with the member libraries and gather feedback.
- Publish MACS as LOD (after obtaining permission from the libraries)
- Enrich MACS with links to DBPedia/Wikipedia by harvesting existing links in the LOD cloud.
- Update the TEL ingestion systems with the enriched MACS, therefore further enriching the TEL Open Dataset

The European Library — Connecting knowledge