Closing the circle: using linked data to enrich authority records and make them linkable on the web

Roxana Popistașu
National Library of Luxembourg
IT department

Stefanie Zutter
National Library of Luxembourg
Cataloguing & Indexing
## National Library of Luxembourg (BnL)

### Legal missions

<table>
<thead>
<tr>
<th>Library Type</th>
<th>Mission</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritage library</strong></td>
<td>- collect, catalogue, preserve, <em>enrich</em> the national heritage</td>
</tr>
<tr>
<td><strong>Scientific and research library</strong></td>
<td>- collect, catalogue, preserve, <em>enrich</em> non-Luxemburgensia collections</td>
</tr>
<tr>
<td><strong>Provide widespread access</strong></td>
<td>- loans and remote consultation</td>
</tr>
<tr>
<td><strong>Manage the network</strong></td>
<td>- manage the Union catalogue and the library network</td>
</tr>
<tr>
<td><strong>Contribute to development</strong></td>
<td>- contribute to the development of library science nationally and internationally</td>
</tr>
</tbody>
</table>

*IGeLU Conference*

Oxford, September 2014
Linking BnL Data...

• Internally
Linking BnL Data...

- Nationally
Linking BnL Data...

- Internationally

IGeLU Conference
Oxford, September 2014
Linking authors (LiDa Project)

Create links between the authors in Aleph (bibliographic & authority records) and their corresponding authors in Autorenlexikon, an encyclopaedia of Luxembourgish literary authors.

Aleph bibliographic records <-> Primo

Aleph authority records <-> Autorenlexikon
LiDa Project

Project partners

National Library of Luxembourg
- Manages ALEPH & PRIMO for the network of libraries in Luxembourg

Centre National de Littérature (National Centre for Literature)
- Manages the content of Autorenlexikon

Magic moving pixel
- IT management side for Autorenlexikon
LiDa Project - Phase 1

• Create links between the authors in Aleph and their corresponding authors in *Autorenlexikon* using *contextual string matching* (author names + associated bibliographic information)

• Not possible (yet) to use authority records
LiDa Project - Phase 1

More info:

• Presentation at IGeLU 2013 in Berlin: Linked Data: From ALEPH/PRIMO to the Dictionary of Luxembourgish Authors and back

• ELCommons Linked Open Data Use Cases and Scenario: Two-way links between authors in Primo & Aleph and Autorenlexikon, an encyclopaedia of Luxembourgish literary authors
LiDa Project - Phase 2

• Create links between the authors in the Aleph authority file and their corresponding authors in Autorenlexikon using existing links

• Part of a bigger project – Authority file:
  • “clean up” existing authority file
  • enrich existing authority records
  • create new authority records
Authority file project

Project areas

- **Organisational**
  - Establish a collaboration framework between BnL and experts from other Luxembourgish organisations

- **Cataloguing**
  - Create new cataloguing template in line with international standards and suitable for a VIAF export

- **Technical**
  - Develop tools for validating and enriching authority records
The librarian’s view

Build initiatives for an enhanced catalog and discovery layer

- Encourage patron engagement with collection
- Promote Luxembourgish content
- Leader in digital content and user experience
Authority data strategy (1)

Harvest data from a range of processes

- legal deposit
- subject experts
- external databases
Authority data strategy (2)

• Make data fit for purpose
• Make data interoperable
  – IDSMARC to MARC21
• Enable semantic web by FRBR-ising authority data
  – AACR2 to RDA
Authority data strategy (3)

- Publish authority data widely
  - BnL discovery tool
  - international aggregators
Authority data management processes

Harvest
- Select data sources for enrichment

Curate
- Transform to standardised & rich data set

Publish
- Expose through existing web services

Start small, evolve step by step
Harvest

Prerequisites

- Find the right data source
- Gain buy-in
- Establish framework for trans-institutional collaboration
- Proof Of Concept for documenting policies and procedures fit for implementation
Data sources

Selection criteria

• Copyright free through legal mechanisms such as contracts, waivers, and licenses (CC-0)
• Data quality
• Data privacy issues
• Freedom of Information Act?
Curate (1)

• Dealing with unstructured information
  ➢ Use data profiling methods

• Nature and scale of the project
  ➢ Document and stabilize evolutionary steps
Curate (2)

• Analyse current state of the database
• Define future requirements for authority data
• Design new authority data schema
• Write system specifications to meet new requirements
• Design governance model
Data governance model

• Provenance indicators
• Plan data maintenance lifecycles with creation/editing policies and procedures
• Keep licensing CC-0 to provision for reuse
• Protection of privacy & confidentiality
• Policies for production of derivatives
• Produce documentation on policies and practices used to create these records
AACR2 authority record

Unique string of characters which represents a particular authority to be traced to bibliographic records.
# FRBR-ised data scheme

<table>
<thead>
<tr>
<th>Leader</th>
<th>0</th>
<th>008</th>
<th>040</th>
<th>100</th>
<th>400</th>
<th>670</th>
<th>046</th>
<th>370</th>
<th>372</th>
<th>374</th>
<th>375</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix pos data</td>
<td></td>
<td>1</td>
<td></td>
<td>Prüm, André</td>
<td>Prum, André</td>
<td>Site Internet uni.lu</td>
<td>Curriculum vitae, publications</td>
<td>19610806</td>
<td>Bridel</td>
<td>Luxembourg</td>
<td>Paris</td>
</tr>
<tr>
<td>Origine</td>
<td></td>
<td>a</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ved nom personne</td>
<td>2</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR personne</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naissance-mort</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieu associé</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activité</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profession</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genre</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Record status**
- **Access point of the entity**
- **Attributes of the entity**
### MARC21 authority record

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader</td>
<td>LDR 001-22-4500</td>
</tr>
<tr>
<td>Fix pos data</td>
<td>008 131230-12-10-30</td>
</tr>
<tr>
<td>Origin</td>
<td>040 LU-CELI</td>
</tr>
<tr>
<td>H-Personal name</td>
<td>100 1a Gleit, Maria 1909-1981</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>SF perso name</td>
<td>400 1a Hofmann-Gleitmann, Hertha</td>
</tr>
<tr>
<td>SF perso name</td>
<td>400 1a Mgl.</td>
</tr>
<tr>
<td>SF perso name</td>
<td>400 1a Dictionnaire des auteurs luxembourgeois</td>
</tr>
<tr>
<td>Data source</td>
<td>670 1a 10990228 19810709</td>
</tr>
<tr>
<td>Assoc dates</td>
<td>046 1a Dictionnaire des auteurs luxembourgeois</td>
</tr>
<tr>
<td>Assoc places</td>
<td>370 1a Crimittschau</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>372 1a Journalistin 1938 1939</td>
</tr>
<tr>
<td>Occupation</td>
<td>374 1a Redakteurin</td>
</tr>
<tr>
<td>Gender</td>
<td>375 1a féminin</td>
</tr>
</tbody>
</table>
Publish

- Publish authority data widely
  - BnL discovery tool
  - International aggregators
Authority file project

Project areas

- **Organisational**
  - Establish a collaboration framework between BnL and experts from other Luxembourgish organisations

- **Cataloguing**
  - Create new cataloguing template in line with international standards and suitable for a VIAF export

- **Technical**
  - Develop tools for validating and enriching authority records
Match authors from the authority file in Aleph and Autorenlexikon using

- existing Aleph links between bibliographic and authority records (system numbers + author headings)

- links created by LiDa between authors bibliographic records and Autorenlexikon (author headings + Autorenlexikon ID)
Create new links (transitivity)
Enrich existing authority records

- Add new information from *Autorenlexikon* to corresponding Aleph fields
- Make sure existing headings are kept → no links to bibliographic records are suppressed
- Correct (some) existing fields
- Validate enrichment of records
- Create Aleph sequential file & Load authority records in Aleph (manage_18, ue_08)
Enrich existing authority records

- Author
  - Autorenlexikon

- LiDa & Aleph

- Author
  - Aleph Authority record

- Aleph

- Author
  - Aleph Bibliographic record 1

- Aleph

- Author
  - Aleph Bibliographic record 2

- Aleph

- Author
  - Aleph Bibliographic record 3

- Aleph

- Author
  - Aleph Bibliographic record 4

- Aleph

- Author
  - Aleph Bibliographic record 5

- Aleph

- Author
  - Auth / Bib record

- existing link

- enriched record
Load new authority records

- Map *Autorenlexikon* information to Marc21 authority record fields
- Create Aleph sequential file
- Load Authority Records in Aleph (manage_18)
- Create links between the new authority records and existing bibliographic records (ue_08 with “N” / manage_102, manage_02, manage_103)
Load new authority records
Live demo

- Links in Primo
- Links in Aleph
- Links in *Autorenlexikon*
- LiDa validation interface
Authority file – next steps

• Luxembourgish authorities
  • Enrich authority file in the production environment (after validation on the development server) – November 2014
  • Export authority file to VIAF – end 2014
  • Add VIAF IDs to Aleph and Autorenlexikon → matching based on IDs – end 2014
  • Collaborate with other Luxembourgish organisations – 2015

• Non Luxembourgish authorities
  • Correct & enrich records using VIAF feedback file
Linked data – new project ideas

• Link authority file authors to:
  • authors in ORBi\textsuperscript{lu} (digital repository of the University of Luxembourg)
  • authors in DigiTool (digitised Luxembourgish newspapers, manuscripts etc.)
  • authors in the new digital preservation system (electronic legal deposit, digital & digitized material)

• Link geographic names in Aleph, DigiTool etc. to GeoNames

• Create links to Wikipedia (DBPedia, WikiData)
Questions, remarks, suggestions?

Thank you!

Roxana Popistașu
IT department
Roxana.Popistasu@bnl.etat.lu

Stefanie Zutter
Cataloguing & Indexing
Stefanie.Zutter@bnl.etat.lu