Identification of entities in the Linked Data collection “Rainis and Aspazija” (RunA)

Anita Goldberga, Maira Kreislere, Anita Rašmane, Aiva Stūrmane,
Bibliography Institute of the National Library of Latvia
Fiesole, 09.05.2017.
Main aims of RunA pilot-project:

• to develop a LD resource for digital humanities to include as many types of digital objects as possible;

• to publish corresponding NLL bibliographic, authority and other metadata on the web;

• to test the co-operation capabilities with other memory institutions
Poets and opinion leaders of the National awakening

Their mutual correspondence has been enclosed in the Latvian national register of UNESCO “Memory of the World” programme
Objectives

Worldwide sources of authority data

Entities
- Time
- Institution
- Work
- Event
- Object
- Person
- Concept

etc.

etc.

etc.

etc.
RunA limitations and findings:

• the scope of the NLL authority database is not sufficient:
  – lack of Work/Expression authority data;
  – “black hole”;
  – abstract entities/subjects;

• drawbacks in identification and linking of Time-span/Event and Place entities (RunA timeline);

• the same types of key entities are identified in the object metadata and in annotated textual documents.
RDA: “To identify and to relate entities”
(Bianchini, Guerrini, 2016)
The way of identification of entities in textual documents

Aspazija\(^1\) in Riga\(^2\)
1894, around September 23(October 5)\(^3\)

I\(^4\) will be off to Liepāja\(^5\) in a few days time, as I am invited kindly by local Society\(^6\) to a “Lost Rights”\(^7\) performance\(^8\); want to see Liepāja\(^9\) also. I will depart on Sunday\(^10\) and be back on Monday\(^11\).

On Ascension\(^12\) day\(^13\), as You\(^14\) know, performed\(^15\) was also “Lost Rights”\(^16\). …
Potential of RDA for RunA (strengths):

- allows to build an entity relationship network including reciprocal relations between entities identified with global URI (Open Metadata Registry);

- language independent relationship designators in RDA Registry, e.g., has founder – P50029;

- supports:
  - the use of other de facto controlled vocabularies, ontologies;
  - hierarchy of entities
Potential of RDA for RunA (weaknesses)

- problems with relationship designator labels in Latvian (conjugations);
  
  700 1# |i Contained in (work): |a Rainis, Jānis, |d 1865-1929 |4 aut |t Dagdas romāns : piecas skiču burtnīcas. [Dagda novel : five sketch books.]

  700 1# |i Ir daļa no: |4 rdaw:P10019 |a Rainis, Jānis, |d 1865-1929 |4 aut |t Dagdas romāns : piecas skiču burtnīcas.

- relations between entities are strictly defined; RunA needs an additional relationship “is mentioned”;
- still lacks the depth that the cross-domains possess
Identification of abstract entities

http://runa.lnb.lv/LNC10-000048138
http://www.yso.fi/onto/koko/p31274
RDA key entities and their relationships with LRM, FRBR and BIBFRAME
Upon evaluating new data models and standards, the RunA project team came to a conclusion:

- the only data standard, the elements/entities of which are identified with globally recognized URI and used by the NLL, is RDA;

- RDA provides tools to change current situation where the NLL data are the final destination for our users
• RunA open data are published as Linked Data (RDF/XML and Turtle)

• Ongoing work:
  – development of special annotation tool for researchers to mark key entities in textual documents;
  – separate data store for entities including all related contextual information and unique identifiers
Acknowledgements:

Dr. comp. Uldis Bojars,
Semantic Technologies
Expert of the NLL

Viktorija Moskina,
Senior Specialist
of the International
Cooperation of the NLL
“Endures one who transforms”

(Rainis)

Thank you!