

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

Anita Goldberga,^(a) Maira Kreislere,^(b)

Anita Rašmane,^(c) Aiva Stūrmane,^(d) Edgars Salna,^(e)

a) National Library of Latvia, Director of the Bibliography Institute b) National Library of Latvia, Data Model Expert
c) National Library of Latvia, System Librarian d) National Library of Latvia, Head of the Data and Knowledge Management Department
e) Mg.Sc.Ing. Datorzinību Centrs, Latvia

Contact: Anita Goldberga, anita.goldberga@lnb.lv.

Received: 10 October 2017; **Accepted:** 31 October 2017; **First Published:** 15 January 2018

ABSTRACT

This paper offers an insight into the work in progress — development of the Linked Data collection RunA <http://runa.lnb.lv/> on two Latvian poets and opinion leaders (Rainis — Jānis Pliekšāns, 1865-1929 and Aspazija, Johanna Emīlija Lizete Rozenberga, 1865-1943) at the National Library of Latvia (NLL). Authors report findings from the theoretical study on identification of RunA collection entities and reality. As collection’s textual materials comprise things and objects of real life, drawbacks in comprehension of NLL authority data are evaluated and perspective with URI identification is emphasised. The authors also address the need to deal with persistent identification for abstract entities. Considering that RunA collection comprises a lot of textual materials (unstructured data) which are to be mutually linked by annotating and identifying key entities (works, persons, institutions, concepts, places, events, etc.), the RDA rules from RunA collection point of view are examined. The main challenges which the creators of RunA faced during pilot project are presented along with corresponding solutions. Conclusions on possible RDA development influence on named entity identification in RunA are reported.

KEYWORDS

EURIG; RDA; Linked Data; Authority data; Identification of entities.

ACKNOWLEDGMENT

This paper was supported by the European Regional Development Fund under the project “IT Competence Centre” (1.2.1.1/16/A/007) and is part of the individual research no. 2.1. “Semantic annotation of textual data in web environment for related data sets.

Dr. comp. Uldis Bojars, Semantic Technologies Expert of the NLL, and Viktorija Moskina, Senior Specialist of the International Cooperation of the NLL

CITATION

Goldberga, A., M. Kreislere, A. Rašmane, and A. Stūrmane. “Identification of entities in the Linked Data collection “Rainis and Aspazija” (RunA)”. *JLIS.it* 9, 1 (January 2018): 83-106 doi: [10.4403/jlis.it-12444](https://doi.org/10.4403/jlis.it-12444).

Preface

This paper presents the work in progress — development of Linked Data collection **RunA (Rainis — Jānis Pliekšāns (1865-1929) and Aspazija, Johanna Emīlija Lizete Rozenberga, (1865-1943))** at the National Library of Latvia (NLL). In June 2016 the pilot collection RunA was opened to public as a virtual collection.¹

Why Aspazija and Rainis? In 2015 Latvia celebrated the 150th anniversary of both poets and opinion leaders of the National awakening that has been included in the UNESCO calendar of events for 2014-2015. The poets' mutual correspondence covering 35 years has also been enclosed in the Latvian national register of UNESCO "Memory of the World" program² to introduce the international audience to some of the highlights of both great personalities. The correspondence covers essential stages of Latvian history and cultural development, as well as realisation of the national state idea.

Rainis and Aspazija's heritage is placed in libraries, archives, museums and private collections inside and outside Latvia (Switzerland, Lithuania, Russia). The initial collection (approximately 500 objects) provides an opportunity to navigate through works of both poets (34 of Aspazija and 49 of Rainis) linking them with other textual, visual and audiovisual objects of the collection: small part of their correspondence, archival documents, photos, posters, audio and video files.

RunA discloses the diversity of both poets' heritage. It offers possibilities that have not been used in Latvia before to discover and to expose values from memory institutions and private repositories as an interlinked web collection. Users of this collection can navigate from specific objects to related historical events (timeline of poets' works and related events) as well as to external objects: related persons, institutions, places, works, etc. (visualization of links between objects; object link graph, i.e. entity "relationship tree").

The pilot collection was created with three main aims in mind. The first is to develop a novel resource for digital humanities to include as many types of digital objects as possible. The second one is to test in practice co-operation capabilities with other memory institutions, such as the National Archives of Latvia, the Institute of Literature, Folklore and Art of the University of Latvia, the Association of Memorial Museums and the Literature and Music Museum. The third one is to publish corresponding NLL bibliographic, authority and other metadata on the Web as Linked Data.

Besides that, RunA is an experimental collection for putting the following new approaches into practice: evaluating new library and web standards; re-using bibliographic and authority data in Web environment, linking them with entities from other present metadata sets and textual documents (literary work annotations and annotated transcripts of correspondence including scientific comments). These textual materials contain a big amount of named entities like people, places, events, etc. that are referenced in bibliographic and authority data. This is an attempt to extract these entities with their attributes and relations from MARC data and link them with marked entities from texts and other type of metadata in Dublin Core (DC).

¹ <http://runa.lnb.lv/>

² <http://www.atmina.unesco.lv/page/79>

Findings from the theoretical study on identification of RunA pilot collection entities

The above mentioned co-operating institutions contributed specific kinds of materials they possess - metadata (authority and bibliographic records), digital copies of letters from Rainis and Aspazija, short descriptions of their creative works (digital texts), legal documents and documents related to education (pictures), transcribed, translated and commented texts of letters, as well as photos, posters, audio and video recordings (digital content complemented with metadata). Bibliographic metadata ensure links to digitized first issues of works by both poets: books and articles kept in the Digital Library of the NLL.

The key materials for resource linking by means of annotation are commented texts of letters and work descriptions. Both are structured (by form - both are texts with clearly distinguishable structure) and freely linguistically expressed by content. Only texts in Latvian are annotated, and only one instance of entity was marked in each separate text document. This decision was made keeping in mind one of main objectives of an annotation - the enrichment of plain text data documents with semantically related data by profiting from unambiguous data records held in authority database and other data sources with persistent identifiers.

Another objective of adding an annotation was to identify and mark entities mentioned in texts, such as persons, organizations, events, places, works and concepts, which have at least one of the following data sources to link to: authority or bibliographic record, descriptive data source or full digital representation. A primary source for metadata is the information from MARC21 databases of the NLL: authority database (people, organizations, events, places, concepts), Union Catalogue (manifestations), National bibliography database (manifestations). Links to the Virtual International Authority File (VIAF) for persons, organizations, places, works, and International Standard Name Identifier registry (ISNI) for persons, institutions are also recorded. Wikipedia in multiple languages was used as a descriptive data source. Full digital representations are gathered through links from the NLL Digital Object Management system (DOM) and the Digital Library of NLL (Book Portal and Portal of Periodicals), The European Library and Project Gutenberg. The NLL portal "In search of lost Latvia" provides links to photos of persons and places, including architecture.

The effort ensures that all abovementioned resources have a unique and persistent URIs. For this reason, letters and descriptions of works are hosted in the NLL DOM system, the NLL bibliographic and authority records are imported in RunA project portal. Import of bibliographic and authority records into the project portal is necessary, because they are originally maintained by library information system ALEPH, which allows deleting database records as a normal routine in main catalogue and other databases. On the contrary, NLL DOM system does not allow deleting objects, so the RunA portal can make use of its URIs.

The annotation process highlights various problems even with the most well-known entity class "person". The biggest problem is that the authority data in libraries include mostly persons related to literature. There is a huge "black hole" where scientists, artists, philosophers, socially and politically influential people, and the great part of other significant (or simply mentioned in letters) persons are "lost". If there is a name without an authority record to link to, sometimes other sources, for example,

a Wikipedia article, genealogical record or a digital photo could be found, but sometimes it is not possible to provide any link.

Entity class “event” proved itself to be a complex issue. It is not a great problem to link world level events, such as the World War I, which has the subject record in the authority database of the NLL and detailed Wikipedia articles in multiple languages. More local and ordinary events, significant perhaps only to correspondents, as this is in the case of Rainis and Aspazija, are mentioned only in letters themselves. These cases highlight the need for marking simple entities “who”, “what”, “where” and “when” with the aim to construct a complex entity - “event”. “Who” gives an active participant of an event, for example, a name of the theatre, where a play is staged. “What” is a name of play (entity class “work”), “where” indicates a place, and “when” is a date. All components together mark an event.

One of the great challenges is creating an annotation of entity “work”. Personal correspondence of two writers refers to poems, dramas, novels, articles, monographs, etc., but also staged plays, movies and songs are mentioned. For example, Aspazija in her letter to Rainis on the 23rd of September, 1894, wrote: “Now I will visit Liepaja for a few days. I’m invited to attend “Zaudētas tiesības” staged there by a charitable society.” “Zaudētas tiesības” (Lost Rights) is a drama by Aspazija, but there is a staged play with identical title mentioned. Sometimes the publication year and the name of the publisher indicate an RDA *Manifestation*, which can then be linked to the appropriate bibliographic record. Mostly RDA *Works* and incidentally *Expressions* are mentioned. The question is: what must be considered as an appropriate link for those if the data in the NLL are not yet FRBRized?

This issue highlights the necessity to create authority records for RDA *Works* and *Expressions* that would be very much appreciated. This question pinpoints the problem of user’s expectations when reading annotated texts with links to other resources. Whether the access to authority or bibliographic record should be given to users instead of the digital resource itself, when it is possible? The pilot collection RunA offers a compromise solution — the webpage that was created stores pictures of book covers, bibliographic records and links to the digitized versions of books held in the Digital Library of the NLL. Sometimes only a link to a public digital library is given, when there is no copy of a certain book in the NLL.

Another kind of problems arise when “work” mentioned in the text is, for example, an article, a poem or a short story. None of them have been published separately — they are parts of greater resources such as magazines, newspapers or almanacs. As a result, they do not have bibliographic records to link to. In many cases they have URIs in full-text digital libraries instead, which could be used to point users to the direct page in a periodical or book.

The situation with non-literary works is different and reveals even more difficulties. Movies, lectures, musical compositions and plays, even buildings, monuments and other architectural creations are works with authors and other responsible persons and organizations involved. In most cases these works do not have bibliographic or authority records at all. Occasionally some photo or musical recording could be found. Purely locally known persons involved do not have any authority records in library world and even in Wikipedia. These entities remain unlinked, because there is no URI to point to.

Challenges and limitations of the pilot collection

The pilot collection development team has identified several types of challenges for organisation and identification of bibliographic entities during the pilot project:

- The scope of the NLL authority database is not sufficient for heritage projects of cross-organizations as it does not cover all kind of essential entities;
- As the NLL catalogue and databases are not yet FRBRized, there is a lack of *Work/Expression* authority records;
- Drawbacks in identification and linking of Time-span/Event entities and Place entities with geographic location/co-ordinates. For example, these entities are not used to create RunA timeline;
- Limitations in identification and linking of abstract entities/subjects;
- The same types of key entities (works, expressions, manifestations; agents: persons, institutions; concepts, places, events) are identified in the object metadata of the collection and also in annotated textual documents. Therefore the necessity to use novel methods to identify and reflect simple or hierarchical relationships between different entities is realized.

The limitations of the pilot project caused by technical solutions are reflected in the article by U. Bojars (Bojars 2016, 21-26).

The main problem related to the identification of entities was that information on entities has been organised in MS Excel tables (instead of a separate database) classifying:

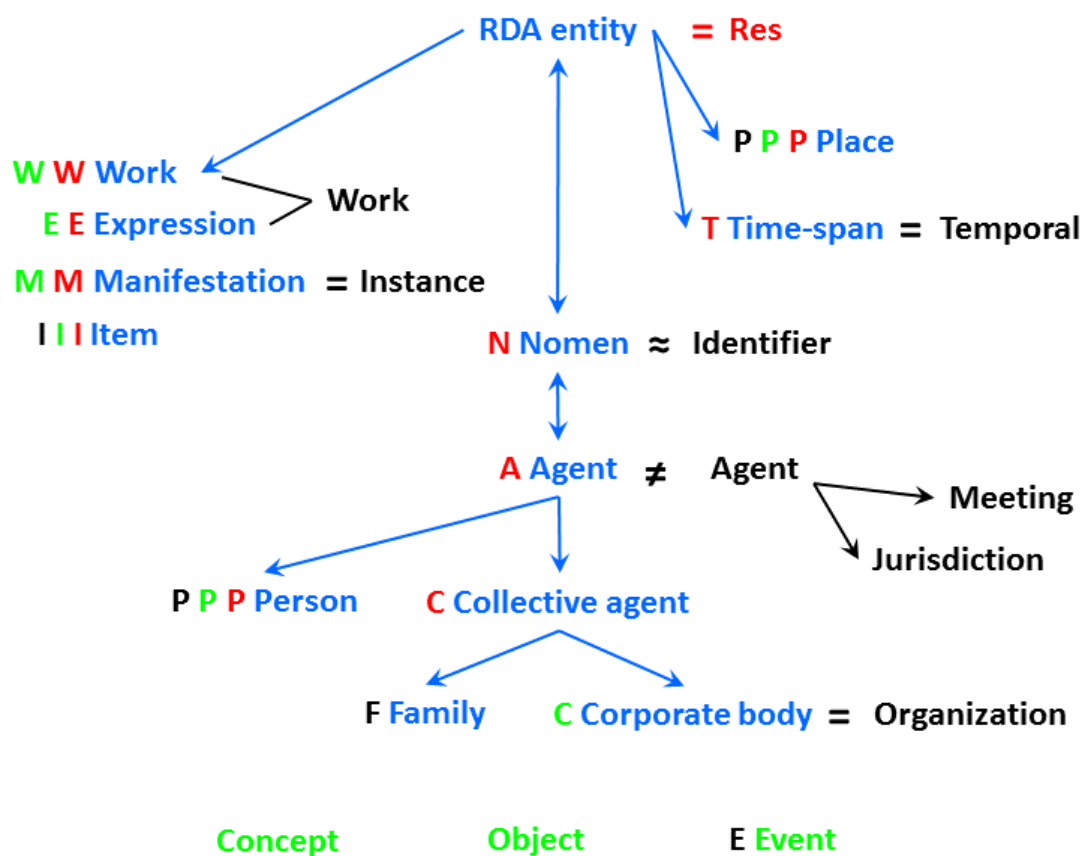
- entity type (class);
- preferred form of entity (from the NLL authority and bibliographic data as well as free chosen forms in Latvian);
- identification of internal (the NLL database or DOM system) or external (VIAF, Wikipedia, etc.) sources.

Identification of external entities/resources is still problematic. In the current development stage mostly local URIs are provided to identify entities.

RunA main groups of entities and the RDA rules

RDA contains all the entities that are of key importance in RunA collection. These named entities are serving as access points in Semantic Web applications, identifying and connecting related entities. “To identify and to relate entities” are two functions realized by RDA (Bianchini and Guerrini 2016, 88).

Some research has been done by the RunA team to compare data models FRBR and IFLA LRM with RDA and BIBFRAME.



RDA LRM Bibframe FRBR

Figure 1. RDA key entities and their relationships with FRBR, IFLA LRM and BIBFRAME.

The issues related to RDA *Work*, *Expression* and *Manifestation* level entity identification and linking are of key importance, for instance, to distinguish between different usage of entity “Faust” (by Goethe) in the correspondence of Rainis and Aspazija. When exactly do both mean “the intellectual or artistic creation”, “the physical embodiment of an expression of a work” or just “Faust” as the literary hero? For example, Aspazija in her letter to Rainis in the middle of June, 1897, wrote: “I will edit Faust as much as I can, and will send it to you tomorrow.” “Faust” mentioned here is a German to Latvian translation of Goethe’s “Faust” by Rainis and Aspazija. Therefore, this is an FRBR *expression*.

RunA project uses RDA defined entities — *Works* (works written by Rainis un Aspazija, their translated works, works that are inspirational sources; related works, which include their works; works by other authors that are based on literary works by Rainis and Aspazija), *Persons* (persons related to Rainis and Aspazija; real, historic or fictional persons mentioned in their works; literary and mythological characters, like Undine), *Institutions* (publishing houses, commercial organizations, cultural and educational institutions), *Concepts* (specific example: bylina — Russian epic poem; abstract

example: nirvana), *Events* (local example: Latvia, German occupation (1941-1944); global example: the First World War), *Places* (geographical — the Baltic sea; historical — USSR, Kievan Rus’).

Named entities in RunA annotations are connected to bibliographic data created in the NLL, and can be identified using authority data. Identifiers are local authoritative data set numbers, VIAF identifiers and ISNI numbers for people and organizations. Authority data are created for *Persons*, *Families*, *Institutions*, *Places*, *Events*, *Concepts*. As was concluded from the pilot project, a huge setback for the project is the fact that authority data for *Works* and *Expressions* have still not been created. It means that it is currently not possible to identify *Works*, their *Expressions*, or to show their interrelations and related *Works* and *Expressions*. Difficulties with person identification, in turn, are often based on the fact that personal names tend to correlate to personal names, works or place names, which are not identified. Collection, for instance, includes several annotations for the name “Aspazija”, which is identified as different persons (the poetess Aspazija and the Ancient Greek feminist Aspazija), as well as with some *Work* titles (a play “Aspazija” by a poetess and a playwright Aspazija, which is based on the novel “Aspasia” by Robert Hamerling). There is also a staged play “Aspazija”, based on the play “Aspazija” written by Aspazija, which can be considered as another *Expression* of this *Work*. If RDA record for the play “Aspazija” is made, it would have to be connected to the *Work* record for the novel “Aspasia” written by Robert Hamerling, by explaining their relations as interconnected works with an attribute “inspired by “Aspasia” — a novel by Robert Hamerling” and “Inspiration for “Aspazija” — a play by Aspazija”. It has to be concluded, that even though it is not possible to show reciprocal relationships between related *Works* in MARC bibliographic records, it can still be done by writing the name of the author and the title of his *Work* in subfield |i together with an appropriate indicator within field 700 in MARC21.

```
100 0# |a Aspazija, |d 1865-1943 |4 aut
```

```
245 10 |a Aspazija : |b Sen-Helladas drāma / |c Aspazija.
```

```
700 1# |i Inspired by: |a Hamerling, Robert, |d 1830-1889. |t Aspasia.
```

```
100 1# |a Hamerling, Robert, |d 1830-1889 |4 aut
```

```
245 10 |a Aspasia : |b novel / |c Robert Hamerling.
```

```
700 1# |i Inspiration for: |a Aspazija, |d 1865-1943. |t Aspazija.
```

This structure still does not create Linked Data, and is only capable of demonstrating relations between the two *Works* visually within a MARC record. In a similar way, MARC environment is not suitable for demonstrating relationships between *Work* and its multiple *Expressions* and *Manifestations* within the structure of Linked Data. A good example of that being the collection of poems by Rainis — “Ave Sol” (*Work*). It is translated into a number of other languages which count as different *Expressions* of this *Work* and only a preferred title and an authority record for author connect their bibliographic records, with *Work* level absent.

100 1# |a Rainis, Jānis, |d 1865-1929 |4 aut

240 10 |a Ave sol! |l Lietuviešu valodā.

100 1# |a Rainis, Jānis, |d 1865-1929 |4 aut

240 10 |a Ave sol! |l Krievu valodā.

It is also difficult to show relationships of this *Work* with related entities like *inspired works* (oratorio, symphony, song cycle) and *CorporateBody* (chamber choir “Ave Sol”) in the MARC structure.

Relationship hierarchy of entities in RunA and RDA

Problems tend to arise upon trying to identify the *Place* entity, when the name of the place coincides with other entities (e.g. personal name). The city of Dagda (Dagda, Dagda municipality) shares the name with one of characters in the *works* of Rainis. If *Person* Dagda does not have its own authority record, it is not a simple task to identify him. Dagda as a fictional character is connected to the series of *works* by Rainis — “Dagda’s sketch books”, which includes a number of poems, each having an individual title. Without *Work* authority record it is difficult to identify separate parts of this *Aggregate Work*. It is possible, though, to show these relations inside the bibliographic records in MARC environment with the help of field 700, where the name of the author, the title and the relationship designator label “(Part of (Work))” is specified under the subfield |i.

100 1# |a Rainis, Jānis, |d 1865-1929 |4 aut

245 10 |a Addio bella : |b no Dagda skiču burtnīcas / |c J. Rainis.

700 1# |i Contained in (work): |a Rainis, Jānis, |d 1865-1929 |t Dagdas romāns : piecas skiču burtnīcas.

100 1# |a Rainis, Jānis, |d 1865-1929 |4 aut

245 10 |a Dagdas romāns : |b piecas skiču burtnīcas / |c J. Rainis ; [Niklāva Strunkes grafiskais ietērps].

700 1# |i Container of (work): |a Rainis, Jānis, |d 1865-1929 |t Addio bella.

By using RDA model, one *Work* record that includes the following information about contained works in general would be necessary: author Rainis, Jānis, 1864-1929, title “Dagda’s sketch books” (alternative title — “The novel of Dagda”), *work* completion *date*, *place* or places where the literary work was written. Furthermore *Work records* for every poem is necessary as well, since all of them can be considered as separate *Works (Aggregate)*. Currently, MARC and ALEPH system allow creating one record with *Work*, *Expression*, *Manifestation* and *Item* attributes included, and not a separate bibliographic record for each FRBR *WEMI* level. It is not possible to demonstrate neither hierarchical levels within the RDA model, nor related entity relations as a Linked Data model. Although by using

RIMMF — an RDA training tool, it is possible to create an individual entry for each *WEMI* level and link them together to show relationships between those levels and entities included.

Element Label	Text	RDA Rule	AAP
Work			
RIMMF identifier °	mairakreislere00000053	no rule	
Authorized access point °	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925)	6.27.1+	
Title of the work		6.2	
Preferred title for the work	Dagdas romāns	6.2.2	<input checked="" type="checkbox"/>
Variant title for the work	Dagdas piecas skiču burtnīcas	6.2.3	
Form of work	Poem	6.3	<input type="checkbox"/>
Date of work	1920-1925	6.4	<input checked="" type="checkbox"/>
Author ×	Rainis, Jānis, 1865-1929 <mairakreislere00000054>	19.2+	<input checked="" type="checkbox"/>
Variant access point °	Dagdas piecas skiču burtnīcas	6.27.4	
Status of identification	fully established	5.7.1	
Expression of work	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925). Text. Latvian. 1963 <mairakreislere00000055>	17.5+	
Container of (work) ×	Rainis, Jānis, 1865-1929. Addio bella <mairakreislere00000058>	25.1+	
Container of (work) ×	Rainis, Jānis, 1865-1929. Čūsku vārdi <mairakreislere00000061>	25.1+	
Container of (work) ×	Rainis, Jānis, 1865-1929. Uz mājām <mairakreislere00000066>	25.1+	
Container of (work) ×	Rainis, Jānis, 1865-1929. Sudrabota gaisma <mairakreislere00000069>	25.1+	
Container of (work) ×	Rainis, Jānis, 1865-1929. Mēness meitiņa <mairakreislere00000073>	25.1+	

Figure 2. RDA relationship hierarchy for the *Work* in parts “Rainis, Jānis, 1865-1929. Dagdas romāns : piecas skiču burtnīcas — Dagda novel : five sketch books (1920-1925)”, for one *Expression* and one *Manifestation*.

Element Label	Text	RDA Rule	AAP
Expression			
RIMMF identifier °	mairakreislere00000055	no rule	
Authorized access point °	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925). Text. Latvian. 1963	6.27.3	
Content type	Text	6.9	<input checked="" type="checkbox"/>
Date of expression	1963	6.10+	<input checked="" type="checkbox"/>
Language of expression	Latvian	6.11	<input checked="" type="checkbox"/>
Summarization of the content	Addio bella ; Čūsku vārdi ; Uz mājām! ; Sudrabota gaisma ; Mēness meitiņa.	7.10	
Language of the content	latviešu	7.12	
Illustrative content	illustrations	7.15	
Work expressed	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925) <mairakreislere00000053>	17.6+	<input checked="" type="checkbox"/>
Manifestation of expression	Dagdas romāns : piecas skiču burtnīcas. Ziemeļblāzma (Jānis Abučs), 1963. Volume <mairakreislere00000057>	17.9+	
Illustrator ×	Strunke, Niklāvs, 1894-1966 <mairakreislere00000056>	20.2+	

Element Label	Text	RDA Rule	AAP
Manifestation			
RIMMF identifier °	mairakreislere00000057	no rule	
Composite key °	Dagdas romāns : piecas skiču burtnīcas. Ziemeļblāzma (Jānis Abučs), 1963. Volume		
Title		2.3	
Title proper	Dagdas romāns	2.3.2	<input checked="" type="checkbox"/>
Other title information	piecas skiču burtnīcas	2.3.4	<input checked="" type="checkbox"/>
Statement of responsibility		2.4	
Statement of responsibility relating to title proper	J. Rainis ; Niklāva Strunkes grafiskais ietērps.	2.4.2	
Publication statement		2.8	
Place of publication	[Vesterosa] :	2.8.2	<input type="checkbox"/>
Publisher's name	Ziemeļblāzma (Jānis Abučs),	2.8.4	<input checked="" type="checkbox"/>
Date of publication	1963.	2.8.6	<input checked="" type="checkbox"/>
Series statement		2.12	
Title proper of series	Raksti /	2.12.2	<input type="checkbox"/>
Statement of responsibility relating to series	J. Rainis ;	2.12.6	
Numbering within series	14. sējums	2.12.9	<input type="checkbox"/>
Mode of issuance	single unit	2.13	<input type="checkbox"/>
Media type	unmediated	3.2	<input type="checkbox"/>
Carrier type	volume	3.3	<input checked="" type="checkbox"/>
Extent statement †	382 pages	3.4	<input type="checkbox"/>
Number of units †	382	3.4.1.3	
Type of unit (text) ×	pages	3.4.5	<input type="checkbox"/>
Base material	paper	3.6	<input type="checkbox"/>
Expression manifested	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925). Text. Latvian. 1963 <mairakreislere00000055>	17.1+	

Relationship	Heading
Work	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925)
Container of (work)	Rainis, Jānis, 1865-1929. Addio bela
Container of (work)	Rainis, Jānis, 1865-1929. Čūsku vārdi
Container of (work)	Rainis, Jānis, 1865-1929. Uz mājām
Container of (work)	Rainis, Jānis, 1865-1929. Sudrabota gaisma
Container of (work)	Rainis, Jānis, 1865-1929. Mēness meitiņa
Expression of work	Rainis, Jānis, 1865-1929. Dagdas romāns (1920-1925). Text. Latvian. 1963
Author	Rainis, Jānis, 1865-1929

Unfortunately, records created in RIMMF cannot be exported to Aleph system. Names of Geographical locations can have multiple meanings and connect to *Works*, *Persons*, *Institutions* and *Events*. For instance, project annotations connect river Daugava to the poem “Daugava” by Rainis, which in turn is connected to a specific historical *Time span* — National awakening, liberation wars and early formation of Latvian state (1915–1919). A play with the title “Daugava” based on this poem was staged later (*derivative work*). M. Brauns, inspired by the poem, wrote a song “Sun, thunder, Daugava” (*inspired work*), which links directly to the *event* “Song festival”. Daugava (location — a river) is also linked to the abstract notion, which only Latvian public can comprehend — the “River of fate”.

Some mutually related entities are still left unidentified and the relationship hierarchy between these elements is yet to be discovered. A play “Fire and night” by Rainis can be called as a bright example of that. Rainis used an epic poem by Andrejs Pumpurs “Lāčplēsis” (“Bearslayer”) as the base for his work “Fire and night”. Folk tales and legends about Lāčplēsis in turn inspired the work “Lāčplēsis” by Pumpurs. This exact work of Rainis is thus a *derived and inspired work*. Relationships between *Works* can be demonstrated using RDA entity hierarchy model.

Relationship	Heading
Work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904)
Inspired by	Pumpurs, Andrejs, 1841-1902. Lāčplēsis (1872-1887)
Expression of work	Pumpurs, Andrejs, 1841-1902. Lāčplēsis (1872-1887). Text. Latvian. 1888
Author	Pumpurs, Andrejs, 1841-1902
Opera adaptation of (work)	Mediņš, Jānis, 1890-1966. Uguns un nakts (1921)
Expression of work	Mediņš, Jānis, 1890-1966. Uguns un nakts (1921). Notated music
Composer	Mediņš, Jānis, 1890-1966
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. Latvian. 1907
Manifestation of expression	Uguns un nakts. Domas. 1907. Volume
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. Belarusian. 1988
Manifestation of expression	Агонь і ноч. Юнацтва. 1988. Volume
Translator	Сёмуха, Васіль Сяргеевіч, 1936-
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. English. 1981
Manifestation of expression	Fire and Night. Echo Publ. West. 1981. Volume
Translator	Millers, Reinholds, 1945-
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. Russian. 1920
Manifestation of expression	Огонь и ночь. Уфимское губернское агенство "Центропечать". 1920. Volume
Translator	Grunts, Jānis, 1892-1950
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. Latvian. 1979
Manifestation of expression	Uguns un nakts. 5. izdevums. Liesma. 1979. Volume
Illustrator	Lamsters, Andris, 1942-
Expression of work	Rainis, Jānis, 1865-1929. Uguns un nakts (1904). Text. Latvian. 2014
Manifestation of expression	Uguns un nakts. Publicētava. 2014. Online resource
Author	Rainis, Jānis, 1865-1929

Figure 3. Relationship tree for the Work “Rainis, Jānis, 1865-1929. Uguns un nakts - Fire and night (1904).”

NLL authority data and RunA entities

Currently movement from authority control in the direction of identity management can be traced in the sphere of library data development. This change is accompanied by the philosophy change and change in the scope of authority data. “Large-scale systems will require a shared store for linked data triples with facilities for Uniform Resource Identifier (URI) management to evolve authority control into entity control” (Dunsire, Fritz and Fritz 2016, 4).

In the case of RunA, information on entities has initially been organised in MS Excel tables. This is just a short-term solution, while an implementation of a special data store for information on entities, with a feasibility to re-use it in other applications is in progress.

“Since 2000 the NLL has maintained the authority records database. It serves as a national bibliographic resource for personal names, corporate bodies, geographic names, and subjects. In summer 2013 authority records for personal names and corporate bodies created by the NLL were successfully incorporated in the Virtual International Authority File (VIAF)” (Goldberga et. al. 2014, 215).

VIAF publishes authority data of the NLL as Linked Data, secures it being linked to other countries’ authority data, Wikipedia, and other resources.

“From the April 2014 on, the NLL uses new RDA MARC21 fields 370 to 378 to create new and enrich existing authority records for persons, corporate bodies and family names. Along with the implementation of the new RDA MARC21 fields and participation in the VIAF, MARC21 field 024 (Other Standard Identifier) and MARC21 field 035 (System Control Number) are added to the NLL authority records, providing International Standard Name Identifier (ISNI) and the VIAF ID” (Ibid, 215–216).

At the beginning of 2017, the NLL authority database had approximately 185,000 authority records for persons and institutions with VIAF ID attached, and about 135,000 of them with ISNI.

Starting from 2016, the NLL has been creating authority data for persons, institutions, families, and geographic names, which fully match current RDA rules; therefore authority records contain much more detailed information related to entities, such as places, activity fields, occupation/profession, gender, language, historical data and other attributes to identify an entity. In accordance with RDA standard authority data allow to reveal relations between persons, families, and institutions, as well as person, institution, and family relations with other FRBR entities — *Work*, *Manifestation* or *Expression*, that match the requirements of Linked Data and Semantic Web.

No authority data for *Works* have yet been created at the NLL, therefore the experience of other libraries and projects on the extraction of information on *Work* entity from bibliographic data was being researched in order to understand the automated FRBRization possibilities and entity matching principles. Here are some examples: National Library of Spain — datos.bne.es; National Library of France — data.bnf.fr; Syrtis (Switzerland), etc.


At the beginning of 2017, mapping was prepared to extract *Work* authority data from the NLL’s bibliographic data (MARC21).

The future plans of the NLL include the creation of authority data collection of national importance that would use the principles of Linked Data. Although at present, navigation from our authority records to VIAF data clusters (by using VIAF ID connected to authority data) is the only way to make our data a part of Semantic Web.

Authority data in RunA collection were chosen to be the basis for identification of named entities during the annotation process, as well as the crossing points for interlinked information based on semantics.

For example, one of the annotations in RunA collection is dedicated to the collection of poems “Tie, kas neaizmirst” (Those who don’t forget, 1911), and it provides the information about some of the mottos of foreign authors included in the edition. One of the authors, whose motto is used in the collection, is Mazzini, Giuseppe (1805-1872) — Italian politician, journalist, activist for the unification of Italy. In this annotation he was identified using the NLL person authority record. While the authority data of the NLL help to identify this person to those who are not familiar with Latvian language, as it is written in the original transcription with the addition of years of life, only the VIAF ID secures the connection to other related information.

Jānis Rainis "Tie, kas neaizmirst" (1911). Anotācija



Tips Teksts
Valoda latviešu
URI <http://dom.lnb.lv/data/obj/64417>
Autors Crule, Astrīda, 1948-
Origināla radīšanas datums 2015
Priekšmets Rainis, Jānis, 1865-1929, Kopsavilkumi
Saite Tee, kas neaizmirst : rudens dzeesma - Rainis, Jānis, 1865-1929 - [1911]
Objekta saite <http://runa.lnb.lv/64417>

RDF/XML **Turtle RDF**

Saistītie objekti

- Rainis, Jānis, 1865-1929
 - Šveice
 - Kastanjola (Lugāno, Šveice)
 - Tičīno (Šveice)
 - Krievija Vēsture
 - Revolūcija, 1905-1907
 - Latvija
 - Dzirciemnieki (Izdevniecība)
- Byron, George Gordon, barons, 1788-1824
- Pellico, Silvio, 1789-1854
- Michelangelo Buonarroti, 1475-1564
- Carducci, Giosuè, 1835-1907
- Mazzini, Giuseppe, 1805-1872
- Rousseau, Jean-Jacques, 1712-1778
- Tee, kas neaizmirst : rudens dzeesma

Datu tīklojums

Objekta teksts

Rainis "Tie, kas neaizmirst" (1911)

Dzejolu krājums "Tie, kas neaizmirst" sacerēts no 1908. līdz 1910. gadam Šveicē, Kastanjolā vai "pa ceļojumiem Tičīnas kantona kalnājos, tādēļ dzejolās tik daudz kalnu ainu un kalnu tēlojumu," skaidro Rainis. Ir beidzies reakcijas laiks pēc 1905. gada notikumiem, un Latvijā ir atsākusies ikdienas materiālā dzīve. Krājums iznāk "Dzirciemnieku" izdevniecībā 1911. gada sākumā. Tas atspoguļo emigrantu izjūtas, un Rainis to nosauc par "Rudens dziesmu". "Ir bijuši un būs vienmēr tie, kas neaizmirst. Tie ir dzīvā vēsture, kas atmiņu dzeļ. Kad dzimtene aizmirst savu nākotni, tie to neaizmirsts, - tie ir dzīvā nākotne, kas dzīvu dzeļ. Un tie paši ir bijuši un būs vienmēr tie, kurus dzimtene izraidīs no sevis, jo viņi nesīs garīgu dzīvību, kādu vēl nespēj dzimtene dzīvot, bet bez kuras tomēr viņai nau nākotnes. Tie ir tie paši, kuri zaudē tēviju un atrod pasauli, zaudē sevi un atrod cilvēci..." Dzejnieks raksta dzejolu krājuma priekšvārda. Ipatnēju koloītu krājumam piešķir daudzā moto, kas ņemti no citātu dzejnieku - Dž. Bairaņa, S. Pellico, B. Mikelandželo, Dž. Karduči, Dž. Madzīni, Z. Z. Rusa darbiem. Nodāļa "Fatamorgāna" uzbūvē talas dzimtenes ainu. Šeit iekļaujas pazīstamais dzejolis "Manu jaunu dienu zeme" gan latviešu literārajā, gan latgaliēšu valodā. Dzeļā organiski ievijas tautas dziesmas - latviešu, lietuviešu, igauņu. Visprecīzāk trimdnieku sajūtas pauž armēņu dziesmas rindas:

"Es izplatots kā tekošs ūdens.
 Es eju no zemes uz zemi."

Par trimdnieku sajūtām liecina nodāļu nosaukumi - "Gaļa bez cēla", "Mērķi aiz mākoniem". Noslēdzošajā nodālā "Kalna galā" Rainis runā par kapu kalnā. *Sansalvatores* virsotnē ir apglābāts polu emigrants, kurš vēlēies no augstā kalna sarudzēt savu dzimteni. Rainis vērsās pie jaunatnes: "Bet tagad mana rudens dziesma jau izdziedāta. [..] Pēc rudens man jāiet atkal jaunā pavasarī. Kas nāks līdzi?" Dzejnieks ir gribējis atgādāt - "kas pagātni piemin, zin nākotni". Tas varētu noderēt visos laikos.

Figure 4. RunA page for description of the work of Rainis — collected poems "Tie, kas neaizmirst" (They, who don't forget). Description mentions Giuseppe Mazzini.

Mazzini, Giuseppe, 1805-1872

Datu tīklojums

ALEPH ID LNC10-000059157
Objekta saite <http://runa.lnb.lv/140593>
Url https://en.wikipedia.org/wiki/Giuseppe_Mazzini

024 ISNI 0000000120984420
100 Persona Mazzini, Giuseppe, 1805-1872
400 Arī: Madzīni, Džuzepe, 1805-1872
400 Arī: Маццини, Джузеппе, 1805-1872
856 Saite VIAF ID <http://viaf.org/viaf/2498020>

RDF/XML **Turtle RDF**

Objekti kuros ir norāde

- Jānis Rainis "Tie, kas neaizmirst" (1911). Anotācija
 - Rainis, Jānis, 1865-1929
 - Šveice
 - Kastanjola (Lugāno, Šveice)
 - Tičīno (Šveice)
 - Krievija Vēsture
 - Revolūcija, 1905-1907
 - Latvija
 - Dzirciemnieki (Izdevniecība)
 - Byron, George Gordon, barons, 1788-1824
 - Pellico, Silvio, 1789-1854
 - Michelangelo Buonarroti, 1475-1564
 - Carducci, Giosuè, 1835-1907
 - Mazzini, Giuseppe, 1805-1872
 - Rousseau, Jean-Jacques, 1712-1778
 - Tee, kas neaizmirst : rudens dzeesma

Figure 5. RunA page for authority record of Giuseppe Mazzini (1805-1872).

Mazzini, Giuseppe, 1805-1872

Mazzini, Giuseppe

1872-1805, **مازینی, جوزیپ**

1805-1872, **מציני, ג'וזפה**

Mazzini, Giuseppe (Genoese revolutionary, 1805-1872)

Giuseppe Mazzini patriota, politico e filosofo italiano

1872-1805, **مازینی, يوسف**

Мадзини, Д. 1805-1872 Джузеппе

VIAF ID: 2498020 (Personal)
Permalink: <http://viaf.org/viaf/2498020>
ISNI: 0000 0001 2098 4420

Preferred Forms

- 100 0 _#a Giuseppe Mazzini #c patriota, politico e filosofo italiano
- 200 _1 #a Mazzini #b , Giuseppe #f <1805-1872>
- 200 _| #a Mazzini #b Giuseppe #f 1805-1872
- 100 1 _#a Mazzini, Giuseppe
- NII 100 1 _#a Mazzini, Giuseppe
- B 100 1 _#a Mazzini, Giuseppe
- LI 100 1 _#a Mazzini, Giuseppe #d (1805-1872),
- QW 100 1 _#a Mazzini, Giuseppe #d 1805-1872
- RD 100 1 _#a Mazzini, Giuseppe #d 1805-1872
- SI 100 1 _#a Mazzini, Giuseppe #g Genoese revolutionary, 1805-1872
- NO 100 1 _#a Mazzini, Giuseppe, #d 1805-1872
- Y 100 1 _#a Mazzini, Giuseppe, #d 1805-1872
- US 100 1 _#a Mazzini, Giuseppe, #d 1805-1872
- DE 100 1 _#a Mazzini, Giuseppe, #d 1805-1872

Figure 6. VIAF page for authority record of Giuseppe Mazzini (1805-1872).

Giuseppe Mazzini

Da Wikipedia, l'enciclopedia libera.

↻ *Disambiguazione* – "Mazzini" rimanda qui. Se stai cercando altri significati, vedi **Mazzini (disambigua)**.

↻ *Disambiguazione* – Se stai cercando l'imprenditore e politico omonimo, vedi **Giuseppe Mazzini (1883-1961)**.

« Ebbi a lottare con il più grande dei soldati, Napoleone. Giunsi a mettere d'accordo tra loro imperatori, re e papi. Nessuno mi dette maggiori fastidi di un brigante italiano: magro, pallido, cencioso, ma eloquente come la tempesta, ardente come un apostolo, astuto come un ladro, disinvolto come un commediante, infaticabile come un innamorato, il quale ha nome: Giuseppe Mazzini. »
(Klemens von Metternich, *Memorie*^[1])


Giuseppe Mazzini (Genova, 22 giugno 1805 – Pisa, 10 marzo 1872) è stato un patriota, politico, filosofo e giornalista italiano, nato nell'allora territorio della Repubblica Ligure, annessa da pochi giorni al primo impero francese.

Le sue idee e la sua azione politica contribuirono in maniera decisiva alla nascita dello Stato unitario italiano; le condanne subite in diversi tribunali d'Italia lo costrinsero però alla *latitanza* fino alla morte. Le teorie mazziniane furono di grande importanza nella definizione dei moderni movimenti europei per l'affermazione della democrazia attraverso la forma repubblicana dello Stato.

Indice [nascondi]

- Biografia
 - Famiglia e giovinezza
 - L'attività conspirativa
 - Dopo l'Unità: gli ultimi anni
 - Le traversie della salma
 - Il mausoleo
 - Dubbi sull'affiliazione massonica
- Il pensiero politico
 - Le idee diffuse in Europa all'epoca di Mazzini
 - La concezione mazziniana
 - Dio e Popolo*
 - Giuseppe Mazzini e Gioacchino da Fiore
 - Una "religione civile"
 - Patria e Umanità
 - L'iniziativa italiana
- La funzione della politica
- La questione sociale
- Le cospirazioni e il fallimento dei moti mazziniani
 - La Giovine Italia (1831)
 - Il fallimento del moto in Savoia (1833)
 - Il tentativo d'invasione della Savoia e il moto di Genova (1834)
 - La tempesta del dubbio* (1836)
 - I fratelli Bandiera (1844)
 - La Repubblica Romana (1849)
 - Il moto di Milano (1853) e la sollevazione in Valtellina (1854)
 - La spedizione di Sapri (1857)
 - Il senso dell'impresa
 - L'appoggio a Garibaldi e gli ultimi tentativi
- Controversie
 - Il conflitto con Cavour
 - I timori di Mazzini per la cessione della Sardegna

Giuseppe Mazzini



Triumviro della Repubblica Romana


Durata mandato 29 marzo 1849 – 1º luglio 1849
(con Aurelio Saffi e Carlo Armellini)

Predecessore Aurelio Saliceti
(con Mattia Montecchi e Carlo Armellini)

Successore Aurelio Saliceti
(con Alessandro Calandrelli e Livio Mariani)

Coalizione Giovine Europa

Dati generali

Partito politico  Giovine Italia/Associazione Nazionale Italiana (1831-1853)
Partito d'Azione (1853-1867)

Tendenza politica Republicanesimo Mazzinianesimo

Titolo di studio laurea in giurisprudenza

Università Università degli Studi di Genova

Professione giornalista e avvocato

Firma *Giuseppe Mazzini*

Figure 7. Wikipedia page for Giuseppe Mazzini (1805-1872).

The annotation to the poem of J. Rainis “Ave sol” (Hail the Sun, 1910) the name of which includes an ancient Roman greeting, is also included in the collection. This annotation also provides the information about one of Latvian cultural symbols, the choir “Ave sol” (*Corporate Body*), which carries the name of the poem. This entity was identified with the NLL authority record, and the data included inside help to identify and provide more exhaustive information about the choir, but the VIAF ID secures the link to other related resources.

Authority data for geographical place names created at the NLL have not yet been included to VIAF. The *Place* entity (building address) is often identified as a specific *Corporate Body* (institution), even though different institutions can occupy that building at different time. It would be better to create authority data by treating buildings like objects, while additionally providing information about their

architects, builders, owners, related institutions, their history, and social significance. This information is an essential part of national cultural and historic heritage, which is absent from data at present.

The entity mentioned in several annotations of the collection — the abstract object “Glass Mountain” can be looked at in the context of the play by Rainis “Zelta zirgs” (The Golden Horse), as well as the building of the National Library of Latvia designed by Gunnar Birkerts. The architect Birkerts was inspired by the *Glass Mountain* depicted in a play of Rainis, as the symbol of National awakening. The object *Glass Mountain* and the abstraction of attaining spirituality by “climbing the Glass Mountain” are connected to both the “National Library of Latvia” as an institution, and with the design of the building by Birkerts. Birkert’s project of National Library of Latvia itself can be considered as *Work*, and according to RDA needs to have a *Work* record, which would include an author — Birkerts, Gunārs (*Agent*); title and its variations (National Library of Latvia, Glass Mountain, Castle of Light, Lighthouse, etc.); date and place of creation of this *Work (project)*; connection to other *Works*: “Zelta zirgs” (The Golden Horse) by Rainis, poem by Auseklis “Gaismas pils” (Castle of Light), song by Jazeps Vitols “Gaismas pils” (Castle of Light).

Talking about the entity “Lāčplēsis” (Bearslayer - literary hero, fairy tale character), it is important not only to identify its relations to different *Agents* but also to show the mutual relations between them. It is possible if we are using the environment, which allows to structure data in RDA entity levels (FRBR *WEMI*). RIMFF environment allows creating data structure that matches the structure of Linked Data. It is possible to create separate *Work*, *Expression*, *Manifestation*, and *Item* entries as well as authority records by linking them together to show mutual relations between these entities.

Identification of entities in unstructured data, as in letters and annotations in RunA, is further complicated by the natural polysemy of language inherent in this data, acronyms, the meaning of which is clear only to authors of the letters, the usage of nicknames and signs, as well as inserting the information in other languages. In addition to that, vague expressions related to time or place like ‘three days ago’, ‘this year’, ‘in the sea side’, ‘the Craftsmen’s hall’, etc., and the necessity to involve well educated and knowledgeable specialists on Latvian culture and literature in the process of identification of entities, makes the process even more cumbersome.

Person identification process can also be tricky because of the usage of natural sequence ‘name, surname’ in the letters, instead of ‘surname, name’ sequence of authority records. Correspondence also includes only names, only surnames, other words like ‘uncle’. Therefore involvement of qualified specialist is needed here as well.

Latvian language is one of inflected languages, where grammatical categories of gender, number or conjugation in conjugable parts of speech are expressed with conjugable word endings. Word ending depends on the context, thus is objected to change. The usage of conjugated words in unstructured data also complicates the process of entity identification, as in structured data like authority or bibliographic data words are usually put in Nominative singular or plural.

In the process of identifying concept entities in unstructured data, one should really keep in mind the natural ambiguity of the language and that by using subject authority records for concept identification, to denote a concept in unstructured data, different synonym can be used. For instance, the concept of ‘feminism’, its synonym form can be ‘emancipation of woman’.

In order to serve better for purposes of entity identification for library created authority data, authority forms would have to be supplemented with a larger amount of concepts, place names and synonym forms. Additionally, perhaps, person authority records should include ‘name surname’ form used in natural language. Furthermore, authority entries should be created for Latvian people and institutions not connected directly to literary process, although significant in other contexts.

Potential of RDA and entity identification with URI

Research done by the RunA collection development team reflects some strengths and weaknesses of RDA.

RDA strengths:

- Allows to build an entity relationship network including reciprocal relationships between entities identified with global URI (Open Metadata Registry);
- There are defined URIs for relationship designators in RDA Registry (language independent), e.g., *has founder* — P50029;
- Compatible with FRBR and IFLA Library Reference Model (LRM); developing under the influence of LRM; supports hierarchy of entities;
- Reduces data duplication;
- Is a candidate to become a data content standard for memory institutions;
- Allows data in different formats (MARC, DC, etc.);
- Supports the use of other de facto controlled vocabularies, ontologies.

RDA weaknesses:

- RDA arrives from the existing library standards and still lacks the depth that the cross-domains possess;
- Problems with relationship designator labels in Latvian (impact of conjugations);

Relationship designator label “Contained in (work)” is a phrase designed for user interface. Exact translation in Latvian — “Ir daļa no (darbs)” does not correspond to the correct use of conjugations in Latvian (correct form: “Ir daļa no darba”). From the user perspective only first part of label “Contained in (work)” — “Contained in” is essential. Other part “(work)” is important for structuring and importing data from MARC to other data environments, so it could be coded, for example, by using CURIE (Compact URI) from RDA Registry.

Example:

700 1# li Contained in (**work**): la Rainis, Jānis, ld 1865-1929 |4 aut lt Dagdas romāns : piecas skiču burtnīcas. [Dagda novel : five sketch books.]

700 1# li Ir daļa no: |4 **rdaw:P10019** la Rainis, Jānis, ld 1865-1929 |4 aut lt Dagdas romāns : piecas skiču burtnīcas.

- Problem in re-using other data. Too much subjectivity in the depth/level of detail;

For example: RDA does not specify whether CD provides information on every *Work* it contains or the whole CD in general (depends on the local decision).

- The necessity to define new relationships that could appear/emerge;

It is possible that the necessity to define new relationships can arise. For example: relationship between the entity ‘building’ and entities ‘architect’, ‘constructor’, ‘owner’.

- To reflect entities mentioned in content of all documents RunA team would appreciate additional relationship in RDA “is mentioned” (for instance, another author, work, concept, etc. is mentioned in content of Work).

Comparing to use of subjects in similar cases relationship “is mentioned” would better serve to the aims of RunA because it is not possible always to find terms from controlled vocabularies.

Entities essential to RunA collection and which are the backbone of authority data have already been defined in FRBR: *Work*, *Expression*, *Manifestation*, *Person*, *Institution*, *Concept*, *Event*, and *Place*. FRAD defines one more entity — Family. Upon evaluating global models, standards and in OMR published namespaces that are of interest to the NLL, the project team came to a conclusion that the only data standard, the elements/entities of which are identified with globally recognized URI and used by the NLL (FRBR is not used) is RDA:

FRBR (1998)	RDA (2010)
<i>Work</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1001	<i>Work</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/Work
<i>Expression</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1002	<i>Expression</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/Expression
<i>Manifestation</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1003	<i>Manifestation</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/Manifestation
<i>Person</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1005	<i>Person</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/Person
--	<i>Family</i> (FRAD) http://rdvocab.info/uri/schema/FRBREntitiesRDA/Family
<i>Corporate Body</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1006	<i>Corporate Body</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/CorporateBody
<i>Place</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1010	<i>Place</i> http://rdvocab.info/uri/schema/FRBREntitiesRDA/Place

FRBR (1998)	RDA (2010)
<i>Concept</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1007	<i>Subject</i> http://rdvocab.info/uri/schema/FRBRentitiesRDA/Subject
<i>Event</i> http://iflastandards.info/ns/fr/frbr/frbrer/C1009	<i>Event</i> http://rdvocab.info/uri/schema/FRBRentitiesRDA/Event

Table 1. Key entities in RunA comparing to FRBR and RDA corresponding entities.

Linked Data representations of RDA components in RDA Registry “allow developers to improve the interoperability of data produced by RDA with a wider base of international standards.” (Dunsire 2016, 17).

It is possible to declare a unique RDA URI while creating MARC21 authority data for persons, families, institutions, places, items, and events. These are possible to obtain from vocabulary embedded in Open Metadata Registry and add to authority records in 5XX reference fields by using |4 Relationship code subfield.

100 1# |a Birkerts, Gunars, |d 1925-

500 1# |i Is architect of: |a Latvijas Nacionālās bibliotēkas ēka |4 rdaa:P50196³

Identification of abstract entities of RunA

To move away from isolated data there is a significant need to deal persistently with identification of abstract entities like “feminism”, “women’s rights”. For example, linking Aspazija’s dramatic works entities with related entities in the correspondence to Rainis.

Applying identification, semantic linking and networking of entities in case of unstructured data sets, there is sometimes a necessity to identify and semantically link concept and event entities. Controlled vocabularies for subjects or ontologies would be the best solution to fulfil the “task of navigating in the world of recorded knowledge” (Bianchini and Guerrini 2016, 2).

“Many data elements ... are created by ontologies and controlled vocabulary terms.” (Bianchini and Guerrini 2016, 4).

NLL uses its own created Subject Headings (NLL-SH) in Latvian language for thematic access, which is a list based on the adapted Library of Congress Subject Headings (LCSH) subject indexing language.

“Most of NLL-SH records were adapted from LCSH by translating preferred labels into Latvian while other records, specific to the NLL and Latvia, were introduced without having a matching LCSH concept. The NLL fully complies with the LCSH system structural regularities — hierarchical relationships, associative relationships and syntax. Library of Congress original subject headings are

³ <http://rdaregistry.info/Elements/a/P50196>

included as *See from* references, thus standardizing the translation and providing for international data retrieval” (Stūrmane, Eglīte and Jankevica-Balode 2014, 20-21).

NLL-SH authority data in RunA collection were used to identify different abstract entities, concepts and events, like “feminism, alchemy, dragons, Christmas, World War I, 1914-1918”, in annotations and comments.

Subject authority records secure entity identification of these concepts and events, as they also contain synonym forms of these concepts and events in Latvian language, chronological data for events, as well as concept names in English, which became possible due to Library of Congress original subject headings. While NLL-SH transition to Linked Data has not been completed, further semantic linking is not possible. Different solutions for NLL-SH linking with subject controlled vocabularies in other languages, especially those represented in form of Linked Data, are being considered and researched.

One possible solution for that is through a connection with LCSH by using Library of Congress original subject headings attached to NLL-SH authority data. U. Bojārs, the researcher of the NLL in his paper “Linking Library Data for Quality Improvement and Data Enrichment” states:

“In the case of taxonomies, such as library authority data, linked records from both taxonomies may contain labels in different languages and these records can be enriched by copying labels across datasets, facilitating creation of multilingual taxonomies. Authority data records may contain links, both internal and external, that can be a valuable resource for data enrichment. Once a link between NLL-SH and LCSH records is established, NLL-SH records can be enriched with links to authority records from the National Library of France and the German National Library that are included in the LCSH dataset”. (Bojārs, Žogla and Eglīte 2015, 187).

The NLL is currently in its transitional period from NLL-SH to a subject indexing system similar to FAST (Faceted Application of Subject Terminology), the syntax of which would be simplified and more suitable for the environment of World Wide Web. At this stage form/genre facets are being created. After this transition is over, it will be possible to make a bond both with FAST and corresponding and interconnected LCSH concepts by using English concepts of FAST, which will be added to the NLL subject authority data. Therefore, the system would be able to provide concept identification and multilingualism.

Transition to a system similar to FAST would solve the problem with entity *Time-span* identification and semantic linking, because the NLL currently does not create chronological term authority data and they are not being used as separate access points in bibliographic records as well. Chronological dates as subdivisions appear only in authority records for persons, institutions, topical subjects or events. In future it will be necessary to create Chronological Term authority records corresponding to FAST, which would serve as identifiers for *Time-span* entity and secure its semantic links.

Bearing in mind that concept level identification needs in Linked Data collections will only increase, the NLL experts are considering possible collaborations with developers of Finnish KOKO ontology <http://seco.cs.aalto.fi/ontologies/koko/>. The potential bases for cooperation — English concepts of LCSH used in the NLL authority records.

Aspazija was concerned by her contemporaries as a prominent defender of feminist ideas in Latvia, which found reflection in annotations and commentaries of the collection. To identify the entity for the concept of “feminism” RunA contains NLL-SH authority record, but by linking English concepts of LCSH of this concept with multilingual Finnish KOKO ontology, identification with URI from this ontology, accessibility in multiple languages as well as access to other related resources becomes possible.

events and action > phenomena > societal phenomena > cultural phenomena > feminism objects > abstract objects > cultural objects > ideas (principles) > feminism		
PREFERRED TERM	feminism	

TYPE	KAUNO Concept KITO Concept KULO Concept LIITO Concept VALO Concept YSO Concept	
BROADER CONCEPT	cultural phenomena ideas (principles)	
NARROWER CONCEPTS	ecofeminism radical feminism	
RELATED CONCEPTS	emancipation female perspective feminists woman's status women's literature women's movements women's rights movements women's studies women	
SCOPE NOTE	Feminism refers to political, cultural, and economic movements aimed at establishing greater rights, legal protection for women and/or women's liberation.	
SOURCE	Helecon Yleinen Suomalainen Asiasanasto#feminismi	
IN OTHER LANGUAGES	feminismi	Finnish
	<i>naisasialiike</i>	
	feminism	Swedish
URI	http://www.yso.fi/onto/koko/p31274	
Download this concept:	RDF/XML TURTLE JSON-LD	
CLOSELY MATCHING CONCEPT	Feminism feminism (sv) feminismi (fi)	Library of Congress Subject Headings Allärs - Allmän tesaurus på svenska YSA - General Finnish thesaurus
EXACT MATCH	feminism feminism feminism feminism feminismi (fi) feminismi (fi)	KITO - Kirjallisuudentutkimuksen ontologia KULO - Kulttuurien tutkimuksen ontologia LIITO - Liiketoimintaontologia YSO - General Finnish ontology KAUNO - ontology for fiction VALO - The Finnish Ontology of Photography

Figure 8. KOKO page for concept “feminism”.⁴

⁴ <http://www.yso.fi/onto/koko/p31274>

SYS ID LNC10-000048138

Object link <http://runa.lnb.lv/LNC10-000048138>

URL <https://lv.wikipedia.org/wiki/Feminisms>

150 Subject Feminisms

450 See also: Feminism

450 See also: Sieviešu emancipācija [emancipation of women — LCSH]

450 See also: Sievietes emancipācija [emancipation of woman]

450 See also: Sieviešu kustība [Women's movement — LCSH]

450 See also: Sieviešu atbrīvošanās kustība [Women's liberation movement — LCSH]

Figure 9. RunA page for authority record on “feminism”.

Summary

Possible solutions for organization and identification of RunA entities according to challenges realized during the pilot project:

- As it is not possible to increase the scope of entity/authority control in the NLL, cooperation with other scientific libraries and the National Archives of Latvia has been commenced;
- Automated data extraction process to create *Work/Expression* authority data has been initiated according to mapping conducted by data experts;
- The NLL plans to integrate Linked Data representations of RDA components from Open Metadata Registry in RunA; to identify *TimeSpan/Event* — the need to create chronological terms according to FAST is realized and to identify *Place* — authority data in accordance with LCSH, Wikipedia can be added;
- The NLL plans to use other controlled vocabularies for identification and semantic linking of abstract entities (for instance KOKO) besides the development of NLL-SH and FAST;
- A promising matter to experiment with for the future development of RunA project will be the exploitation of RDA relationship designators marking named entities from different resources and identified entities in textual documents. RunA will additionally have relationship designators with URI in annotations of textual documents;
- A great part of entity classes used in annotations are essentially RDA key elements — *Person, Family, Corporate Body, Place, Event, Subject, Work, Expression, Manifestation*. RDA

defines properties and subproperties (expressing relations) which are identified with URI. This is perfectly suitable for linking entities in collection such as RunA. The only drawback is the “human” part of representing data - RDA properties must be translated in RunA interface languages.

One of RunA collection main tasks is an evaluation of new library and web standards and re-use of bibliographic and authority data in Web environment linking with entities from other metadata sets and objects. Why? Because currently in majority of cases the NLL data are the final destination for our users: they do not allow contextualisation and navigation to full texts, objects or other data sets, i.e. it is not possible to navigate from “strings to things”.

RDA provides memory institutions with novel rules to identify detailed relationships between entities: simple and hierarchical. However it will take a long time to adapt and implement these rules in practice.

Future objective of this collection is to reflect objects, their attributes and relationships using key entity classes, subclasses and the hierarchy of their relationships according to IFLA LRM semantic levels.

RunA Linked Data interface and open datasets are published and perhaps some researchers may find them to be useful.

Ongoing work on development of special annotation tool and separate data store for entities including all related contextual information and unique identifiers is still in-progress.

References

- Bianchini, Carlo, and Guerrini, Mauro. 2016. “RDA: a content standard to ensure the quality of data”. *Italian Journal of Library, Archives and Information Science*. Vol. 7, 2:83-98. Accessed March 17, 2017. DOI: <http://dx.doi.org/10.4403/jlis.it-11709>
- Bianchini, Carlo, and Guerrini, Mauro. 2016. “The international diffusion of RDA: a wide overview on the new guidelines”. *Italian Journal of Library, Archives and Information Science* Vol. 7, 2:1-5. Accessed March 17, 2017. DOI: <http://dx.doi.org/10.4403/jlis.it-11844>
- Bojārs, Uldis. 2016. “Case study: Towards a linked digital collection of Latvian cultural heritage” - In *CEUR Workshop Proceedings*, 1608: 21-26. ISSN 1613-0073
- Bojārs, Uldis, Žogla, Arturs and Eglīte, Elita. 2015. “Linking Library Data for Quality Improvement and Data Enrichment”. In *Proceedings of 4th International Conference on Data Management Technologies and Applications*, edited by Markus Helfert, Andreas Holzinger, Orlando Belo, and Chiara Francalanci. Vol.1: DATA, 184-188. ISBN 978-989-758-103-8
- Dunsire, Gordon, and Fritz, Deborah, and Fritz, Richard. 2016. “Instructions, interfaces, and interoperable data: the RIMMF experience with RDA” (paper presented at: IFLA WLIC 2016 — Columbus, OH — Connections. Collaboration. Community in Session 93 — Cataloguing and Information Technology). Accessed March 16, 2017. <http://library.ifla.org/id/eprint/1324>

Dunsire, Gordon. 2016. "Towards an internationalization of RDA management and development". *Italian Journal of Library, Archives and Information Science* Vol. 7, 2:307-30. Accessed March 17, 2017. DOI: <http://dx.doi.org/10.4403/jlis.it-11708>

Goldberga, Anita, Kreisler, Maira, Sauka, Jogita, Stūrmane, Aiva and Virbule, Inta. 2014. "RDA: From Strategy to Experiments and Implementation in Latvia (Including an Overview of the Situation in the Baltic States)." *Journal of Library Metadata*, Vol. 14, 3-4, 205-21. Accessed March 21, 2017. DOI: <http://dx.doi.org/10.1080/19386389.2014.992710>

Stūrmane, Aiva, Eglīte, Elita, and Jankevica-Balode, Mārīte. 2014. "Subject Metadata Development for Digital Resources in Latvia". *Cataloguing & Classification Quarterly* 52(1): 20-31.