Automatic Publication under Linked Data Paradigm of Library Data

ALIADA, an Open Source Solution to Easily Publish Linked Data of Libraries and Museums

ALIADA Project Consortium
SWIB15, November 23-25, 2015
Hamburg
The challenge: why LOD in LM

- A **global pool of shared data** that can be re-used to describe resources will avoid the redundant effort of the current cataloging processes.

- The **use of the Web** and Web-based identifiers will make up-to-date resource descriptions directly citable by catalogers.

- Linked Data **is more durable and robust** than metadata formats that depend on a particular data structure.

- Developers will also **no longer have to work with library-specific data formats** (MARC, LIDO).

- With Linked Open Data, libraries can **increase their presence on the Web**, where most information seekers may be found.

http://www.w3.org/2005/Incubator/lld/wiki/Benefits
The challenge: how to start

• Cataloguing data according international conceptual models and standards (FRBR, BIBFRAME, CIDOC-CRM, …)

• Exporting records to standard metadata schemes (MARC, LIDO or Dublin Core) and formats (XML)

• Selecting an ontology

• Converting MARC/LIDO/DC metadata to RDF statements

• Linking their own dataset to other datasets (one domain or multidomain)

• Publishing data as 5-star Linked Open Data
“Librarians and curators are experts in cataloguing and making accessible their resources, but the don’t know about Linked Data technology, so they need an ally”
ALIADA, the ally to publish LODLM

- Open source Java application to automatically publish as Linked Data the metadata created by a library or museum management system

- Supported metadata types (types of datasets): bibliographic records, authority records, descriptions of museum objects and other information resources

- Compliant with MARC XML, LIDO XML and Dublin Core formats

- Conversion to RDF triples (mapping) according to the ALIADA ontology, mainly based on FRBRoo, SKOS, WGS84 and Foaf ontologies

- Linking to other datasets, such as Europeana, British National Bibliography, Spanish National Library, Freebase Visual Art, DBpedia, Hungarian National Library, Library of Congress Subject Headings, Lobid, MARC codes list, VIAF Virtual International Authority File or Open Library

- Automatic publication of dumps (URIs) and SPARQL Endpoint on DataHub
ALIADA, the ally to publish LODLM

- EU FP7- ENV-2012 Collaborative project 2013-2015
- Partners: Art museums, libraries, ILS vendors, researchers on Semantic Web technology
- Final release: October 2015
- Open source community expected
- ALIADA is free software, you can redistribute it and/or modify it under the terms of the GNU GPL v3 (License)
ALIADA, the ally to publish LODLM

- M11 – September 2014: 1st Usability workshop
- M12 – October 2014: 1st Prototype + opening to the community
- M15 – January 2015: 1st deployment
- M16 – February 2015: 2nd Usability workshop
- M23 – September 2015: Final Usability workshop
- M24 – October 2015: 2nd Prototype & 2nd deployment + opening to the community
1st prototype (2014)

- User interface in Spanish and English.
- Validation of imported records (MARC Bibliographic and LIDO)
- Mapping templates (FRBRoo)
- RDF-izer ([ALIADA ontology](#))
- Linking to some datasets: Europeana, BNB, BNE, Freebase, Dbpedia, NSZL, Geonames and MARC Code Lists (SPARQL endpoint)
- Linked data server creation + SPARQL endpoint + URIs dereferencing.
- Linked dataset validation: through a number of SPARQL queries.
2nd prototype (2015)

- Translation of DublinCore XML and MARC XML Authorities to ALIADA ontology.
- Validation of RDF dataset consistency.
- NER (Named Entity Recognition) for some text free elements
- Ad-hoc linking to some of the listed external datasets, which do not provide a SPARQL endpoint such as VIAF, LOBID, Open Library and Library of Congress Subject Headings.
- Links disambiguation: the system offers to the user a set of possible ambiguous links: so the user can decide which links are correct and which ones should be rejected.
- Advanced URI de-referencing and creation of a web page for the generated dataset.
- Publication in CKAN of the created linked dataset + DataHub LOD validator.
- Translation of the user interface to Italian and Hungarian.
- ALIADA offers REST services in order to be integrated with other systems: ILS and CMS.
ALIADA, LOD main features
ALIADA, LOD main features

• ALIADA RDFizer
  - Scalability (RESTful application, Apache Camel asynchronous channels, JEE web application)
  - Modularity (Conversion templates are configurable and extensible)
  - Reusability (Standalone installation of the RDFizer)
  - Easy to use/maintain (Conversion job is controlled by the so-called “conversion templates”, which are runtime-interpreted scripts very easy to maintain)
  - Easy to extend (The library is free to create their conversion template, producing an arbitrary output format, with another ontology)
  - Validation before the conversion (Jena OWL Micro Reasoner)
**ALIADA, LOD main features**

- **NER of free text fields.**
  - A dedicated component for doing NLP (Natural Language Processing). Recognizes sequences of words in a text which are the names of things, such as person, company names and places.
  - In LIDO it is applied to `<lido:descriptiveNoteValue xml:lang="en" lido:type ="physical-description">Sculpture of Mozart</lido:descriptiveNoteValue>`
  - In MARC it is applied to
    - "522" "a": Geographic Coverage Note.
    - "525" "a": Supplement Note.
    - "520" "a": Summary, etc.
    - "520" "b": Expansion of summary note.
  - The NER results are stored as RDF triples that enrich the owning records
ALIADA, LOD main features

- Linking to Datasets that provide SPARQL endpoint

### ALIADA dataset

- F3_Manifestation_Product_Type – Title
- E39.Actor
- E21.Person
- F10.Person

### Actor_Appellation

- E18.Physical_Thing – Appellation
- E73.Information_Object – Title
- E58.Place
- F9.Place
- wgs84:lat
- wgs84:long

### Place_Appellation

### E56.Language

### BNB

- dc:title

### Europeana

- edm:ProvidedCHO

### NSZL

- foaf:Person
- ifla-frbr:C1001
- ifla-frbr:P3001 / ifla-frad:P4033
- ifla-frbr:C1005
- ifla-frbr:P3039 / ifla-frad:P4031

### BNE

### Freebase

- book.book
- visual_art.artwork
- film.film

### DBpedia

- Agent:name
- Place:name
- MARC_Country
- MARC_GeographicArea
- MARC:Language
- geo:name
- wgs84:lat
- wgs84:long

### MARCCode Lists
ALIADA, LOD main features

- Linking to Datasets that do not provide SPARQL endpoint

  - E39_Actor
  - E21_Person
  - F10_Person
  - Actor_Appellation
  - Actor_Appellation ecrm:P3_has_note StringToSearchFor

  - F3_Manifestation_Product_Type - Title
  - Title ecrm:P3_has_note StringToSearchFor

  - E39_Actor - Actor_Appellation
  - Actor_Appellation ecrm:P3_has_note StringToSearchFor

  - E89_PropositionalObject ecrm:P129_is_about skos:Concept
    skos:Concept skos:prefLabel StringToSearchFor

  - E1_CRM_Entity P137 exemplifies skos:Concept
    skos:Concept skos:prefLabel StringToSearchFor

- ALIADA dataset

\[
\begin{align*}
\text{http://www.viaf.org/viaf/AutoSuggest?query=} & \text{StringToSearchFor} \\
\text{https://openlibrary.org/search?title=} & \text{StringToSearchFor} \\
\text{http://api.lobid.org/resource?name=} & \text{StringToSearchFor & format=ids} \\
\text{http://api.lobid.org/organisation?name=} & \text{StringToSearchFor & format=ids} \\
\text{http://id.loc.gov/search/?q=} & \text{StringToSearchFor & q=cs:http://id.loc.gov/authorities/subjects&format=xml} \\
\end{align*}
\]

- VIAF
- Open Library
- LOBID: Bib. resources
- LOBID: Libraries & rel. organisations
- Library of Congress Subject Headings
ALIADA, LOD main features

• Links disambiguation

<table>
<thead>
<tr>
<th>Ambiguous links (158)</th>
<th>All links (294)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show 150 entries</td>
<td></td>
</tr>
<tr>
<td>Allida URI</td>
<td>External URI</td>
</tr>
<tr>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td><a href="http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/00412194a-0d92-3af1-ad04-10fc8d137e755">http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/00412194a-0d92-3af1-ad04-10fc8d137e755</a></td>
<td><a href="http://rdf.freebase.com/ns/m.0df605">http://rdf.freebase.com/ns/m.0df605</a></td>
</tr>
<tr>
<td><a href="http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/1087318-7ba7-3af9-ad60-7375fa0e659a">http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/1087318-7ba7-3af9-ad60-7375fa0e659a</a></td>
<td><a href="http://data.nytimes.com/m.058w3">http://data.nytimes.com/m.058w3</a></td>
</tr>
<tr>
<td><a href="http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/26795-55-7977-32b4-8a2-704474a65377">http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/26795-55-7977-32b4-8a2-704474a65377</a></td>
<td><a href="http://rdf.freebase.com/ns/m.000.7">http://rdf.freebase.com/ns/m.000.7</a></td>
</tr>
<tr>
<td><a href="http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/34dc6d1-41e6-4aff-1b7d-a578d0735679">http://alaida.scanbox.net:8891/id/collections/library/bk/211_Person/34dc6d1-41e6-4aff-1b7d-a578d0735679</a></td>
<td><a href="http://rdf.freebase.com/ns/m.06r9f9">http://rdf.freebase.com/ns/m.06r9f9</a></td>
</tr>
</tbody>
</table>

ALIADA. SWIB15 November 23-25, 2015 Hamburg
• Advanced URI de-referencing

URI regulation: http://www.w3.org/TR/cooluris/

• Be on the web
  – Machines and people should be able to retrieve a description about the resource identified by the URI from the Web.
  – Machines should get RDF data and humans should get a readable representation, such as HTML

• Cool URIs
  – Simplisity: Short and mnemonic
  – Stability: Does not change as long as possible
  – Managability: Keep all URIs in a dedicated subdomain
• URI name convention of ALIADA: URI structure

{domain}/{type}/{concept}/{class}/{reference}.{format}

data.szepmuveszeti.hu/id/collections/museum/E18_Physical_Thing/szepmuveszeti.hu_object_29
## URI name convention of ALIADA: URI structure

<table>
<thead>
<tr>
<th>Extension</th>
<th>Media type</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>.html</td>
<td>text/html</td>
<td>HTML</td>
</tr>
<tr>
<td>.json</td>
<td>application/rdf+json</td>
<td>JSON</td>
</tr>
<tr>
<td>.jsonld</td>
<td>application/ld+json</td>
<td>JSON-LD</td>
</tr>
<tr>
<td>.nt</td>
<td>text/plain</td>
<td>N-Triples</td>
</tr>
<tr>
<td>.opac</td>
<td>OPAC of the institution (if restful)</td>
<td>OPAC</td>
</tr>
<tr>
<td>.rdf</td>
<td>application/rdf+xml</td>
<td>RDF/XML</td>
</tr>
<tr>
<td>.ttl</td>
<td>text/rdf+n3</td>
<td>N3/Turtle</td>
</tr>
</tbody>
</table>

http://data.szepmuveszeti.hu/doc/collections/museum/E18_Physical_Thing/szepmuveszeti.hu_object_29.nt
### ALIADA, LOD main features

- **Dataset Default Web Page created by ALIADA**

  ![Artium](image)

  **Aliada dataset**

<table>
<thead>
<tr>
<th>Description</th>
<th>Open linked data from the Library and Museum of ARTIUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>ARTIUM</td>
</tr>
<tr>
<td>Source</td>
<td><a href="http://biblioteca.artium.org">http://biblioteca.artium.org</a></td>
</tr>
<tr>
<td>Created</td>
<td>2015-10-05</td>
</tr>
<tr>
<td>Contributor</td>
<td>Aliada Consortium</td>
</tr>
<tr>
<td>License</td>
<td><a href="http://creativecommons.org/publicdomain/zero/1.0/">http://creativecommons.org/publicdomain/zero/1.0/</a></td>
</tr>
<tr>
<td>Number of triples</td>
<td>11566</td>
</tr>
</tbody>
</table>
ALIADA, LOD main features

- Publication in CKAN Datahub
Data Hub LOD Validator

A handy record validator is provided at [http://validator.lod-cloud.net/validate.php](http://validator.lod-cloud.net/validate.php) to check that at least the minimum required information is present.

After applying such validator over our dataset published by ALIADA, a completeness level of 3 is reached. The following step to reach the top completeness level of 4 is to e-mail to the contacts indicated at [http://lod-cloud.net/](http://lod-cloud.net/) page.
• Integration of ALIADA with the LMS (RESTful API)
ALIADA, LOD main features

Integration of ALIADA with the LMS (RESTful API)
DEMO SITE

Demo video: https://vimeo.com/aliadaproject

ARTIUM’s datasets
http://datos.artium.org/

MFAB’s dataset
http://data.szepmuveszeti.hu/

ARTIUM’s RDF data store (queries)
http://datos.artium.org/sparql

MFAB’s RDF data store (queries)
http://data.szepmuveszeti.hu/sparql
ALIADA open source community

ALIADA site has been updated with the second release, the version resulting from the usability and performance tests as well as from the second deployment in public institutions.

http://www.aliada-project.eu/
The source code of the second release has been uploaded to the GitHub’s ALIADA repository and the technical documentation, the user manual and the wiki have also been updated.

https://github.com/ALIADA/aliada-tool/
THANK YOU FOR YOUR ATTENTION! ANY QUESTIONS?

info@aliada-project.eu