One Supplier’s Approach to BIBFRAME/Linked Data

Tiziana Possemato
Chief Information Officer

ALA Midwinter 2016
BIBFRAME Update Forum Agenda
Current activity and infrastructure

Casalini Libri produces, for publications from Romance language countries, more than 20,000 original bibliographic records in RDA as a member of the Program for Cooperative Cataloguing (PCC)

Bibliographic records are created using the WeCat cataloguing module of the OLISuite ILS (developed by @Cult) in native MARC 21/RDA format

Authority control on names, titles and series headings

Maintenance of the authority databases (NACO and SACO)
The three areas of activities towards the BIBFRAME/Linked Data environment

In order to be ready with the concept of evolution from a web of documents (the traditional web) to a web of data (the semantic web) and with the started transition from MARC formats to Linked Open Data, we are investing in the study and implementation of projects that go in this direction, more recently with emphasis on the model proposed by the Bibliographic Framework Initiative (BIBFRAME).

To this aim, we focus on the following three areas of activities.
The three areas of activities towards the BIBFRAME/Linked Data environment

1. The enrichment of MARC records to simplify BIBFRAME conversion

2. The use of a framework to automate the conversion from MARC to RDF, using BIBFRAME vocabulary

3. The creation of a FRBR/BIBFRAME layer starting from bibliographic and authority records, to help librarians and end users in LOD fruition
1. Enrichment of MARC records to simplify BIBFRAME conversion

Additional MARC tag fulfilment and treatment in order to simplify the conversion into BIBFRAME without losing content: the MARC record is enriched (through manual and automatic processes) with tags and subfields, in particular with the addition of a certain number of local and global identifiers.

This builds the precondition to allow the conversion of MARC into Linked Open Data by any party.
URI Management System

We are improving in the WeCat cataloguing module of OLISuite a «URI Management System», to manage identifiers for each access point or heading.

See as an example in the following slides the authorized access point for Franz Kafka and, in the first column (URI), the number of URIs associated to the heading.
URI Management System (WeCat screen)
The cataloguer can check, modify, delete or add other identifiers to the same heading.
Access points and URIs

The URIs associated to a heading can be used in varying and useful ways.

In the data export/conversion process we can choose how many URIs to make available for each heading, how to associate them to the heading, how to show them in relation to data use and formats.
Access point and URIs (example 1)

As $0$ associated to access point in the MARC bibliographic record:

=LDR 00560nam a2200181 4500
=001 000000127573
=003 CaOOAMICUS
=005 20160108094931.0
=008 160107s\\\\\\\\\\\it\\\\\\\\\\\000\u\ita\r
=040 \$aAtCult$bita
=100 1\$aKafka, Franz,$d1883-1924$0(isni) 0000 0001 2280 370X.
=245 03$aLa metamorfosi /$cFranz Kafka.
=260 \$aMilano :$bLa spiga,$c2002.
=300 \$a61 p.; $c18 cm
=336 \$atext$2rdacomment
=337 \$aunmediated$2rdamedia
=338 \$avolume$2rdacarrier
=997 \$aPS
Access point and URIs (example 2)

As specific tag in the MARC authority record:

=LDR 00698nz 2200145 4500
=001 000000000617
=005 20160108125155.0
=008 751003s1974

=024 7\$a56611857$2viaf
=024 7\$a000000012280370X$2isni

=040 \$aPS$bita
=100 1\$aKafka, Franz$2d1883-1924
=400 1\$aKafka, F.$q(Franz)$d1883-1924
=670 \$aWikipedia, Oct. 25, 2012$bFranz Kafka; born 3 July 1883 in Prague; died 3 June 1924 Kierling near Vienna; an influential German-language writer of novels and short stories, regarded by critics as one of the most influential authors of the 20th century. Kafka was a Modernist and heavily influenced other genres, including existentialism)
Access point and URIs
(example 3)

As RDF property in the triples produced in the conversion process:

\begin{verbatim}
001 000000000617
024 7 $a56611857$2viaf
024 7 $a00000012280370$2isni
100 1 $aKafka, Franz
\end{verbatim}
Access point and URIs
(example 4)

Another example of identifiers used as RDF property of an entity type Person:

001 000000000617
024 7 $a56611857$2 viaf
024 7 $a000000012280370$2 isni
100 1 $aKafka, Franz

<atcult:617-kafka-franz>
<rdf:type>
<bf:Person>

<atcult:eb-617>
<rdf:type>
<bf:Identifier>

<atcult:eb-617>
<bf:local>
<atcult:617-kafka-franz>

<atcult:eb-617>
<bf:identifierValue>“617”

<atcult:617-kafka-franz>
<bf:hasAuthority>
<rdf:resource="http://viaf.org/viaf/56611857”>

<atcult:617-kafka-franz>
<bf:hasAuthority>
<rdf:resource="http://isni-url.oclc.nl/isni/000000012280370”>
2. Use of a framework to automate the conversion from MARC to RDF

The conversion from different formats to RDF is realized within the WeCat cataloguing module, that embeds micro-agents software, each one mapped on a specific MARC tag/subfield in order to convert and export it as Linked Open Data.

The same conversion process can be activated independently from an ILS, using data in different formats (MARC, xml, Lido, etc.).

Automatic conversion in RDF is realized through the ALIADA framework, applying the BIBFRAME vocabulary.
### OliSuite/WeCat: from MARC 21 to RDF

Example of the conversion process activated in the ILS: at the end of the cataloguing workflow, the user can click on the RDF button to convert/publish the record as RDF triples, using the **ALIADA** framework.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tipo</td>
<td>Materiale linguistico - Monografia</td>
</tr>
<tr>
<td>Livelli</td>
<td>Non verificato</td>
</tr>
<tr>
<td>Leader</td>
<td>00585nam a2200217 u 4500</td>
</tr>
<tr>
<td>Nr identificativo</td>
<td>000000087121</td>
</tr>
<tr>
<td>Data ultima variazione</td>
<td>20151231160817.0</td>
</tr>
<tr>
<td>Nr controllo sistema</td>
<td>11062561</td>
</tr>
<tr>
<td>Lingua record</td>
<td>ita</td>
</tr>
<tr>
<td>Fonte catalogazione</td>
<td>PUL(R)</td>
</tr>
<tr>
<td>Accesso persona</td>
<td>Manzoni, Alessandro, 1785-1873.</td>
</tr>
<tr>
<td>Titolo autorizzato</td>
<td>Promessi sposi</td>
</tr>
<tr>
<td>Titolo proprio</td>
<td>Gli sposi promessi / di Alessandro Manzoni.</td>
</tr>
<tr>
<td>Pubbl/dist/stampa</td>
<td>Milano : Bianchi Giovini, 1943.</td>
</tr>
<tr>
<td>Descr. fisica</td>
<td>642 p. : ill. ; 23 cm</td>
</tr>
<tr>
<td>Serie</td>
<td>Aretusa</td>
</tr>
<tr>
<td>Serie - Titolo autorizzato</td>
<td>Aretusa.</td>
</tr>
<tr>
<td>Biblioteca</td>
<td>PUL</td>
</tr>
<tr>
<td>Collocazione</td>
<td>. 206 I 20</td>
</tr>
</tbody>
</table>

**Buttons in the interface:**
- Modifica (F2)
- Duplica (F8)
- Cancellia
- RDF
- Status Copie
- Posseduto
- Crea FRBR
- Scarto
- Opac

**Images in the interface:**
- A screenshot of the cataloguing interface with a highlighted RDF button.
The framework used to convert and publish data in RDF is **ALIADA: Automatic publication under Linked DAta Paradigm of library Data.**

The project is co-financed by the European Union’s Research and Innovation funding programme for 2007-2013 (FP7).

- 5 partners from 3 different countries (**Italy**, **Spain**, **Hungary**)
- 2 IT companies: **@CULT**, **SCANBIT**
- 2 museums: **ARTIUM** (Spain), **Museum of Fine Arts Budapest** (Hungary)
- 1 research institute: **TECNALIA** (Spain)

Project duration: 24 months (from November 2013 to October 2015)
Results available as open-source at [www.aliada-project.eu](http://www.aliada-project.eu)
The conversion process from any format to RDF

**RESOURCES**
- Library Management System (ILS)
- Museum Collection Management System (MMS)
- Content Management System (CMS)

**METADATA CREATORS**
- Librarians
- Curators

**ALIADA**

**LINKED DATA CLOUD**

**BROWSERS**
- Google

**IT COMPANIES**

**OTHER PUBLIC AND CULTURAL INSTITUTIONS**

http://lod-cloud.net/
ALIADA conversion & publishing layers

- Conversion:
  - MARCXML
  - Dublin Core
  - Validation of Input Data
  - ALIADA ontology
  - RDFizer
  - MARCXML2RDF
  - DublinCore2RDF
  - LIDO2RDF
  - Other RDFizer

- Linking:
  - RDF Triple Store
  - Linked Data Server
  - W3C SPARQL endpoint
  - Links Discovery
  - Linked Dataset

- Publication:
  - Validation RDF
  - Linked Dataset
  - Creation CKAN DataHub page

- User Interface:
  - RDFizer
  - Linked Data Server
  - CASALINI LIBRI
Ontologies used in the framework

ALIADA

- BIBFRAME*
- SKOS
- WGS84
- Foaf
- OWLTime

Additional ontologies used in the ALIADA framework:

- FRBRoo (part of the first release)
- DCMI Metadata Terms
- RDF Schema
- RDA elements

*BIBFRAME added in the current release in progress
The asynchronous pipeline

ALIADA building block, realized through Apache Camel. The process is split into atomic pieces (*processors*), each of these responsible for a small part of the overall task. Each processor can act as a *splitter* or *aggregator* and can achieve content manipulation on the incoming message.
It's just an asynchronous pipeline!

The high-level workflow in ALIADA is as follows: before proceeding with the conversion of a record, the pipeline looks up the Work/Person cluster to gather information about a given entity, in order to disambiguate and uniquely identify things in the out-coming dataset.

A set of MARC records go through the pipeline, which splits, processes and converts them.
ALIADA Conversion templates

ALIADA converts each incoming record by means of *Conversion templates*. Each template associates:

- a MARC record belonging to the incoming data-stream
- with a set of (conversion) rules associated with one or more ontologies.

```
001 27283 020 1 $a880921191X

001 27283 100 1 $aCollodi, Carlo.
```
Another example of the conversion process from UniMARC to BIBFRAME

The underlying idea is to provide a “Domain Specific Language” to minimize the learning curve and hopefully allow non-technical people to create their own template quickly and easily.
ALIADA Conversion rules

Technically, a conversion template is a file containing conversion rules, expressed in a high-level programming language.

For instance, the rule:

```
#set ($s = #uri('Work' 1643)
$s $is_a #bf("Work").
```

produces the following:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix bf: <http://bibframe.org/vocab/>

```

The conversion rules can be centralized and then reused, in order to gain speed for the implementation of new rules, e.g. adding more mappings with different ontologies.
3. The creation of a FRBR/BIBFRAME layer from bibliographic and authority records

The existent catalogues are description, above all, of manifestations/instances. We tried to give an answer to the requirement to re-design the data model with a system that derives data from existent records to produce a new Person/Work layer. The process creates for each Person entity a ‘cluster’ of possible variant forms, and does the same for associated Works.

- **Person cluster**: creation of a unique *name access point* for Person names.
- **Work cluster**: each Person is associated to his Work.

Each Work cluster is linked to *Instance* titles.
The loading process and creation of clusters

The loading processor and creation of Person/Work clusters: an important step of the process retrieves data from external authority files, such as VIAF, using the specific APIs.
Cluster makers - Person (example 1)

This chart and the following example show the mechanism for associating names from different records in a single Person cluster.
Cluster makers - Person (example 2)

| 001  | 00002 |
| 200  | $aVan Ness,$bHendrick C. |
| 997  | $aAUTHORITY |

| 001  | 8379 |
| 701  | $aVan Ness,$bHendrick C. |
| 997  | $aUNIBAS |

| 001  | 173506 |
| 701  | $aVan Ness,$bHendrick C. |
| 997  | $aUNINA |

| 001  | 1317 |
| 701  | $aVan Néss$b, Hendrick C. |
| 997  | $aUNISANNIO |

| 001  | 56522 |
| 701  | $aVAN NESS,$bHendrick C |
| 997  | $aUNISA |

**ID cluster:** 1425297  
**Author:** Van_Ness, Hendrick C.  
**Other forms:**  
Van_Ness, Hendrick C.

**ID cluster:** 965648  
**Author:** Van Ness, Hendrick C.  
**VIAF ID:**

**Other forms:**  
VAN NESS, Hendrick C.  
Van Ness, H. C.  
Van Néss, Hendrick C.  
Van Ness, H. C. (Hendrick C.)  
Van Ness, Hendrick Charles
Cluster makers - Person (example 3)

Reading the MARC record we obtain:
• the association of names through a weighted algorithm for comparison
• the identification of already existing clusters or creation of new clusters
• the aggregation of different forms of names through VIAF APIs
Biblioteche

Individuazione

Raccolta

Selezione

Elaborazione

Search Engine

RDF Store

LOD Cloud

Linked Open Services Platform (Elaborazione dati in RDF)

VIAF

Virtual International Authority File

Authority file locali

ALIADA

(Elaborazione dati in RDF)

Search Engine

BIBFRAME-UP

Linked Open Services Platform

LOD Cloud
The BIBFRAME-UP: a three layer architecture

**Person/Works**

- **Wikipedia**: L’enciclopedia libera
- **Sant’Agostino d’Ippona**

**VIAF**

Virtual International Authority File

**Library of Congress Authorities**

**Contra academicos**

- De Beatâ Vitâ
- De civitate Dei contra paganos

**Instances**

- *Contra Academicos*
- *La vita felice*
- *La città di Dio*
- *Über das Glück*

**Item**

- **University of Salerno**
- **Biblioteca Centrale di Ateneo**
- **University of Sannio**
- **L’Orientale University**
- **University of Naples "Parthenope"**
The BIBFRAME-UP Portal: synthesis

1st layer - Person/Work: the set of data related to Person and Work, in RDF, saved in a SPARQL endpoint and made available by specific search and presentation functions.

2nd layer - Instance: bibliographic data indexed in SOLR search engine, that is able to produce new different data aggregations in facets (such as publication date, language, publisher, edition, etc.). This layer provides users with a wide range of search and navigation functions.

3rd layer - Item: holdings data, related to copy information, coming from the local OPAC or local system of each specific library.
Conclusions

Where we are now?

The action plan of the three exposed areas of activities is in progress.

We are in a phase of analysis and development where opportunities to share experiences, doubts and input from the community about expected priorities can be crucial.

Any cooperation and manifestation of interest by institutions and people, to share and disseminate activities and results, is very welcome.
Thank you

Tiziana Possemato
Chief Information Officer

tiziana.possemato@casalini.it
www.casalini.it

ALA Midwinter 2016
BIBFRAME Update Forum Agenda