Implications of BIBFRAME and Linked Data for Libraries & Publishers

Michele Casalini
Chief Executive Officer

Charleston Library Conference
November 3rd, 2016
Casalini’s activity in the bibliographic field

Casalini Libri produces, for publications from Romance language countries, more than 30,000 original bibliographic records in RDA as a member of the Program for Cooperative Cataloguing (PCC)

Bibliographic records are created using the inhouse 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)
# Casalini’s BIBFRAME Conversion, Distribution & Publication Options

<table>
<thead>
<tr>
<th>PRODUCTION TOOLS</th>
<th>SOURCES</th>
<th>INPUT</th>
<th>CONVERSION TOOLS</th>
<th>OUTPUT</th>
<th>DELIVERY &amp; PRESENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>@Cult</td>
<td>ilibri</td>
<td>M21</td>
<td>ALIADA</td>
<td>FRBRoo</td>
<td>CKAN</td>
</tr>
<tr>
<td>Open Library Innovation Suite</td>
<td>Bibliographic Database</td>
<td>DC</td>
<td>Conversion tools</td>
<td>LODIZE</td>
<td>Library’s RDF Store</td>
</tr>
<tr>
<td>WeCat module</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ilibri-Up Portals</td>
</tr>
<tr>
<td>Adempiere ERP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SHARE Platform</td>
</tr>
<tr>
<td>@Cult</td>
<td>casalinilibri</td>
<td>LIDO XML</td>
<td>ALIADA extentions</td>
<td>Linked Data as a Service</td>
<td></td>
</tr>
<tr>
<td>WeCat module</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPORT TOOLS</td>
<td></td>
<td>M21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Via ALIADA REST API protocol</td>
<td></td>
<td>DC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torrossa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Text Platform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Casalini’s BIBFRAME Conversion, Distribution & Publication Options:

- @Cult is the technological partner of Casalini Libri

For further information please contact bibframe@casalini.it

October 2016
Among the library community needs

Transition from the 1960’s encoding and exchange format to the new data model in light of the current technology

Workflow changes within libraries and their partners

Adaptations of the library systems components

Greater library data interoperability and visibility (and “out of the silos”)

Technological preconditions for new cooperative institutional initiatives

Concrete advantages for the end users
How to catch the opportunities of LOD

We do not have yet a new mature scenario, but we are in a unique period of time where standards, models and guidelines are being refined.

At the same time tools are being developed, tested and shaped in order to redesign the new information chain of the coming decade.

Short term projects seem to be the best approach to reach concrete results and to give answers to the open questions necessary to have the basis for solid upcoming plans.
Casalini’s areas of activity towards the BIBFRAME and Linked Data environment

1. The enrichment of MARC records with identifiers (URIs) to simplify the conversion into LOD - BIBFRAME
2. The use of a framework to automate the conversion from MARC to RDF, using the BIBFRAME data model, with the integration of other models then useful
3. The creation of a FRBR/BIBFRAME publication layer starting from bibliographic and authority records, to help librarians and end users in LOD fruition
4. The creation of original data in RDF with implications on the retro-conversion into MARC
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 addition of a number of local and global identifiers.

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

Profiling options are foreseen to handle the single URI source preferences by libraries.
URI Management System

Implementation in the WeCat cataloguing module of OLISuite (@Cult) a «URI Management System», to manage identifiers for each access point or heading.

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

In the data export/conversion process it can be defined how many URIs are available for each heading, how to associate them to the heading, how to show them in relation to the data use and formats.
The cataloguer can check, validate, modify, delete, or add other identifiers to the same heading.
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$2rdaccontent
=337 \$aunmediated$2rdamedia
=338 \$avolume$2rdacarrier
=997 \$aPS
Access point and URIs (example 2)

Example of identifiers used as RDF property of an entity type *Person*:

001 0000000000617
024 7 $a56611857$ viaf
024 7 $a000000012280370$ isni
100 1 $aKafka, Franz
Current implementation stage

In progress:

• Development completion of database enhancements useful to embed the URIs in the headings;

• API developments to interact with the various sources;

• Export tool developments in order to handle libraries’ personalized URI source profiling.
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.
ALIADA: the RDF conversion & publication framework

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 (Casalini’s technological partner), 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
Ontologies implemented in the framework

Additional ontologies used in the ALIADA framework:

- FRBRoo (part of the first release)
- DCMI Metadata Terms
- RDF Schema
- RDA elements
- BIBFRAME 1.0 added with the Spring 2016 release, BIBFRAME 2.0 implemented in October 2016.
The asynchronous pipeline

In the ALIADA building block 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.
Current implementation stage

July 2016:
• BIBFRAME vrs 1 test data released in nt (N-Triples), in ttl (Turtle), in xml (RDF/XML) formats for one month production data of six North American research libraries.

• This data is also available online in the Blazegraph database that supports SPARQL queries at the address http://lod.casalini.it:9999/

October 2016:
• BIBFRAME 2.0 test data released
3. The creation of a FRBR/BIBFRAME layer from bibliographic and authority records

The current catalogues are description, above all, of Manifestations/Instances. The goal has been 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 Works.

Each Work cluster is linked to the **Instance** titles (i.e. the publications); and from there to the **Items** (OPAC level).
The loading process and creation of clusters

The creation of Person/Work clusters is a crucial step of the process that retrieves and reconciliates data relying also on external authority files, such as ISNI, VIAF etc.
Example of automated processes to reconcile data

If the form is found in an Authority record it has the heaviest weight in the reconciliation algorithm.
The controlled name access point

The Person cluster for sant’Agostino:

- Augustinus, Aurelius <354-430> (AUF) (VIAF)
- Augustinus, s., vesc. d’Ippona, 354-430 (VIAF)
- Augustine, Saint, Bishop of Hippo (VIAF)
- Augustinus, Aurelius santo (FED) (BAS)
- Augustinus Hipponensis 354-430 (FED)
- Aurelius Augustinus Hipponensis 354-430 (AUF) (FED) (BAS)
- Agostino d’Ippona santo ; 354-430 (FED)
- Agostino, Aurelio santo ; 354-430 (FED) (BAS)
- Augustinus, Aurelius santo (BAS)
- Augustinus, Aurelius <santo> (ORI) (SAL)
- Augustinus santo (SAN)
- Augustinus, Aurelius (SAN)
- Augustinus : von Hippo santo (SAN)
- Agostino : d’Ippona santo (SAN)
- Agostino santo (SAN)
- Augustinus, Aurelius santo ; 354-430 (BAS)

Name cluster ID 245

VIRTUAL UNION CATALOGUE
(search for Works, or for the Instances level with Name ID cluster 245)
Retrieve Works associated to Person: the *controlled title access point*

The *Titles cluster* for *Varro, Marcus Terentius (ID 69518)*:

- De lingua latina (ID 986)
- Saturae menippeae (ID 1314)
- De vita populi romani ad Atticum (ID 2135)
- De gente populi romani (ID 2136)
- Antiquitatis (ID 2137)
- Logistorika (ID 2138)
- De re rustica (ID 855)
- Res rusticae (FED)
- Économie rurale (SAL) (VIAF) (FED)
- Del Camp (SAL)
- Gespräche über die Landwirtschaft (FED)
- Varro the farmer a selection from the Res rusticae (FED)
- Rerum rusticarum libri tres (VIAF) (BAS)

**Title cluster ID 855**

VIRTUAL UNION CATALOGUE
(search for Instances, or for the Items level with Title ID cluster 855)
SHARE is a common discovery layer developed by @Cult based on Linked Open Data through identification, reconciliation and clustering processes applying the BIBFRAME data model.

The first project went into production in Spring 2016 with seven universities using different ILS: http://catalogo.share-cat.unina.it/sharecat/clusters?l=en

A North American SHARE prototype project is in the preliminary stage.
The SHARE Catalogue project consultation portal is characterised by architecture made up of three different levels:

- **Person - Works**
- **Instances (publications)**
- **Items**
SHARE - Person cluster

Other name forms

- Mises, Richard von <1883-1953>
- Мизес, Р. 1883-1953 Рихард
- Von Mises, Richard
- Mises, R.von 1883-1953 Richard von
- Richard von Mises österreichischer Mathematiker
- Mises, Richard von
- Misès, R. de.
- Mises, Richard von, 1883-1953
- Von Mises, Richard, 1883-1953
- Mises, Richard Martin von, 1883-1953
- Misès, Richard de <1883-1953>
- Von Mises, Richard <1883-1953>
- Mises, Richard von <1883-1953>
- MISES, Richard : von

Results of a Person cluster for the Richard von Mises entity with the different name forms stemming from:
- Authority files
- VIAF
- Referrals (to authority record)
- Forms used in bibliographic records
SHARE - Author data enrichment

This person:
- **Tolstoy, Lev Nikolaevich** 1828-1910
  - ID: 160060

Other name forms:
- Tolstoj, Lev Nikolaević 〈1828-1910〉
- Tolstoj, Lev Nikolaević
- Толстой, Л. Н. 1828-1910 Лев Николаевич
- Tolstoj, Lev
- Толстой, Лев Николаевич, граф, 1828-1910
- Tolstoj, L. N. 1828-1910 Lev Nikolaević
- Tolstoj, L.N. (Lev Nikolaević), 1828-1910
- Tolstoj, Leo (Russian writer, philosopher, 1828-1910)
- Лев Николаевич Толстой русский писатель и мыслитель
- Tolstoj, Lev, 1828-1910

Works:

This person in:
- Library of Congress (data.bnf.fr)
Grouping under a single work title of the many publication titles in the catalogue for *Promessi sposi*.

One work title brings together more than 70 different publications catalogued by the different libraries, which the end user can access with just one search.
Clicking result on the title *Promessi sposi*
Current implementation stage

March 2016:

**SHARE Catalogue**, in production
A virtual union catalogue of seven university libraries
using different ILS, exploiting enriched RDF converted data using the BIBFRAME data model.
It’s the result of a two year project.

May 2016:

**ilibri-up**, prototype
Knowledge base prototype for cluster and URI registry development.

October 2016:

**SHARE Virtual Discovery Environment**, start of a prototype project for a group of North American research libraries with the aim to address the heterogeneity of different systems, habits and cataloguing traditions.
Thank you.

For further information:

Michele Casalini  
michele@casalini.it

Tiziana Possemato  
Chief Information Officer  
tiziana.possemato@casalini.it

The described activities are based on the technological partnership of Casalini Libri and @Cult.