AthenaPlus workshop
Innovative tools and pilots for access to digital cultural heritage in the framework of Europeana and national systems

Experiments of Linked Open Data in the framework of the National Union Catalogue (SBN)

Patrizia Martini (ICCU)
Achille Felicetti (VAST-LAB – PIN)

- ❑ Is the national center which promotes and coordinates cataloguing and digitisation activities of the Italian libraries
- ❑ Is responsible for guiding, producing, adapting and disseminating the standards and rules for cataloguing and digitisation of the library heritage
- ❑ Is in charge to coordinate and to manage important national systems and census: National Library Service (SBN), National Census of Italian XVI century editions (EDIT16), Census of Manuscripts of Italian Libraries (Manus), Internet Culturale, CulturaItalia…
- ❑ Is in charge to coordinate many European and international projects in the field of digital cultural content

http://www.iccu.sbn.it

ICCU institutional tasks

❑ Promotion and production of cataloguing rules and guidelines and diffusion of international standards

  ➢ SBN Cataloguing rules

  ➢ Translation of IFLA international standards and models (ISBD, FRBR…)

  ➢ Translation of RDA rules (LoC, JSC)

❑ Activities in close connections with international bodies: IFLA, PUC, VIAF, EURIG, Dublin Core Metadata Initiative…

Experiments of LOD in SBN

The SBN National Coordination Committee has committed to ICCU a study for the definition of a data model for publishing OPAC SBN data in Linked Open Data

<u>Initial steps:</u>

❑ Analysis of a basic methodology for creating open linked data from the bibliographic data according to the international standards and cataloguing rules in SBN

❑ Drawing the logic scheme relating to bibliographic content, attribution of descriptive, semantic, management metadata

❑ Selection of a first set of bibliographic records and export them from OPAC SBN in UNIMARC format

❑ Development of namespaces to accommodate UNIMARC element sets and value vocabularies following the FRBR family of models

The working group:

Francesco Gandolfi, Patrizia Martini, ElisaSciotti (ICCU), Achille Felicetti (PIN)

Experiments of LOD in SBN

Further steps:

- Data conversion into RDF(Resource Description Framework) encoding syntactic basis of the Semantic Web and linked data.

- Growing an interface that allows libraries to connect to the SBN Index, and derive triples that constitute entities of their interest (work, expression, etc.), converted to standard formats, and afterwards usable within their catalogue

- Creation of links to data sets of international open linked data projects: VIAF, Europeana, Linked Heritage, DBPedia, etc.

- Publication of data in RDF/XML with different modes: with appropriate access URI and URL, through interfaces, SRU, OAI SPARQL Endpoint

- definition of EP-SPARQL search function for published datasets consultation

- validation, ensuring the respect of the formats required by the semantic web, and multiple tests on the data set functionality in a manner consistent with the principles and purposes of the semantic web

Overview

- ❑ Evolution of library Linked Data RDF Description Framework and the Semantic Web
- ❑ Identity management
- ❑ Schema, mapping, interoperability
- ❑ Open publishing
- ❑ Identifiers
- ❑ Universal Bibliographic Control
- ❑ Role of national libraries and bibliographic agencies
- ❑ Bibliographic standards (IFLA)
- ❑ Impact and implications for catalogues and cooperation

http://www.ifla.org/

http://www.gordondunsire.com/presentations.htm

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

ATHENA
PLUS
Access to cultural heritage
networks for Europeana

**The UNIMARC format
towards Linked Open Data**

❑ **IFLA 2013** "Expanding MARC Metadata Services with Linked Open Data"

This session focused on concepts, strategies, experiences, technical issues and prospects of linked open data (LOD) applied to MARC and UNIMARC records.

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

Linked data initiatives

- UDC Consortium: published a selection of around 2,000 UDC classes in 16 languages online as the UDC summary (RDF version in development)

- Virtual International Authority File (VIAF): a set of linked controlled vocabularies (authority records of personal names) by national bibliographic agencies

- ISBD: prescribes vocabulary control for the data in the Area 0 for content form and media type. Terms for the elements (content form, content qualification, and media type) are taken from closed lists

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

SBN

5.000 libraries linked to the SBN Catalogue through 84 nodes

- 12 M **resources**
- 3,5 M authors
- 66 M locations

Daily traffic

- 200.000 daily requests from libraries to Index

- 52 M searches

SBN Catalogue

OPAC SBN

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti
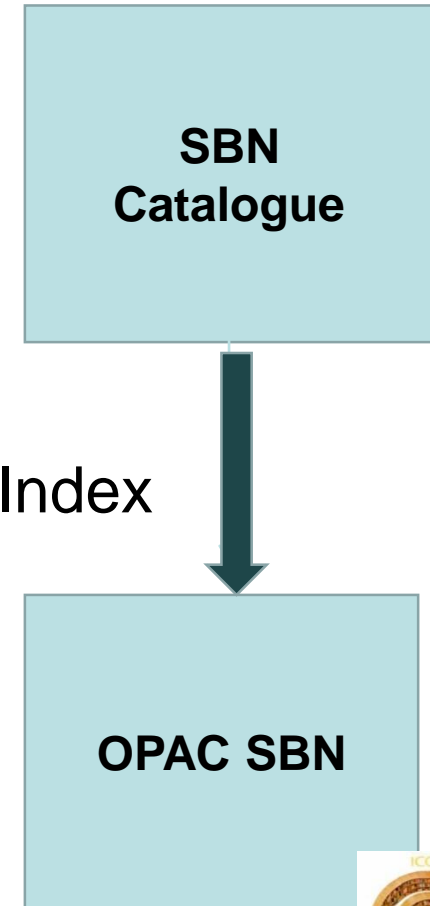
SBNMARC protocol

SBNMARC is a **communication protocol** that simplifies interacting between LMSs and Index, because it uses XML messages and a schema based on **UNIMARC semantic:** concepts and attributes.

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

Cataloguing in SBN

Shared cataloguing is based on:

❑ a «core» of common data

❑ specific data for different types of resources (Music, graphic, cartographic)

❑ authority files:

- Uniform titles
- Names (personal and corporate body names)
- Printers' devices
- Places of publication
- Semantic headings (subjects, classification)

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

Modern: monographs and periodicals post-1830

Antique: monographs until 1830.

Music: printed and manuscript music, including music librettos, musical sound recording, musical videos

Graphic: drawings, paintings, posters, pictures, photographs, etc.

Cartographic: Manuscript and printing cartographic resources

OPAC SBN

The conversion

☐ To select a first set of records of **modern material** ("core elements") and create an element for every allowed combination of tag, indicators, and subfield.

☐ Representation at finest level of granularity to be captured at the level of the UNIMARC subfield code and in the case of value vocabularies, the notation code and corresponding term, to avoid loss of information

☐ Intellectual value of UNIMARC is preserved by a finest-grained semantic representation

Element sets & value vocabularies

- Open Metadata Registry
  - http://metadataregistry.org/
  - FRBR/FRAD/FRSAD family
    - http://iflastandards.info/ns/fr/
  - ISBD
    - http://iflastandards.info/ns/isbd/
  - CIDOC-CRM
    - http://cidoc-crm.org
  - RDA
    - http://rdvocab.info/
- Linked Open Vocabularies
  - http://lov.okfn.org/dataset/lov/

Value vocabulary

- Dewey Decimal Classification

  http://dewey.info/

- Nuovo Soggettario (BNCF)

  http://thes.bncf.firenze.sbn.it/

RDF

- Resource Description Framework

Data format that supports simple, single metadata statements known as triples

- – Each statement is in 3 parts
- Based on description logic
  - – subject-predicate-object statements
- Also specifies relationships between things
  - – thing-relationship-thing statements
    - Can be used for navigating between, or integrating, information from multiple source

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

- FRBR is a conceptual <u>model</u> built on a E-R methodology which is intrinsically applicable to representation in RDF, while ISBD is a <u>data standard</u>
- Design of the RDF representation of ISBD involves:
  - the treatment of aggregated statements in a defined number of elements within the areas;
  - the treatment of mandatory and optional elements and areas;
  - the order of areas and elements within an area;
  - the repeatability of areas and elements;
  - the treatment of punctuation and its double function.

FRBR and CIDOC-CRM

- FRBR/CIDOC CRM Harmonisation
- Expressing the IFLA FRBR model with concepts and tools provided by the CIDOC CRM
- Aligning the two object-oriented models
- Semantic interoperability between library and museum information

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

FRBR and CIDOC-CRM

- **FRBRoo**
- Formal ontology
- To capture and represent the underlying semantics of bibliographic information
- To facilitate the integration, mediation, and interchange of bibliographic and museum information
- Service Providers and Content Providers domains
- Well established and widely used standard
- Systems more stable
- Information more sharable

Tools

- ## W3C Semantic Web tools
  - http://www.w3.org/2001/sw/wiki/Tools
- ## W3C LinkingOpenData
  - http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
- ## Open Metadata Registry (for vocabulary management)
  - http://metadataregistry.org/
- ## W3C SKOS
  - http://www.w3.org/2004/02/skos/

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

RDF Triples

- Metadata expressed as "atomic" statements
  - A simple, single, irreducible statement
    - The title of this book is "Cataloguing is fun!"
- Statements Constructed in 3 parts
  - "Triple"
- The title of this book is "Cataloguing is fun!"
  - Subject of the statement = *Subject*: This book
  - Nature of the statement = *Predicate*: has title
  - Value of the statement = *Object*: "Cataloguing is fun!"
- This book –> has title –> "Cataloguing is fun!"
  - *subject –> predicate -> object*

- Create or obtain URI for the Resource described
- Obtain URI for UNIMARC tag/subfield
  - Direct from tag/indicators/subfield encoding
- Obtain URI of value of subfield, or use a literal value
  - URI from vocabulary or UNIMARC Authority
- Publish triple

Related standards: Other

- <u>The National Library of Sweden</u> has developed a methodology for representing MARC21 records in RDF and implemented it for LIBRIS, the Swedish Union Catalogue

- <u>The Vocabulary Mapping Framework (VMF)</u> project [funded by UK Joint Information Systems Committee (JISC)]:

  - to develop a major expansion of the RDA/ONIX framework for resource categorization

  - to create a tool to support the automated mapping of vocabularies from metadata standards of use to the JISC community, which includes research, teaching, and learning environments

  - CIDOC CRM, FRAD, FRBR, MARC21 and RDA vocabularies [included] & ISBD and UNIMARC [represented]
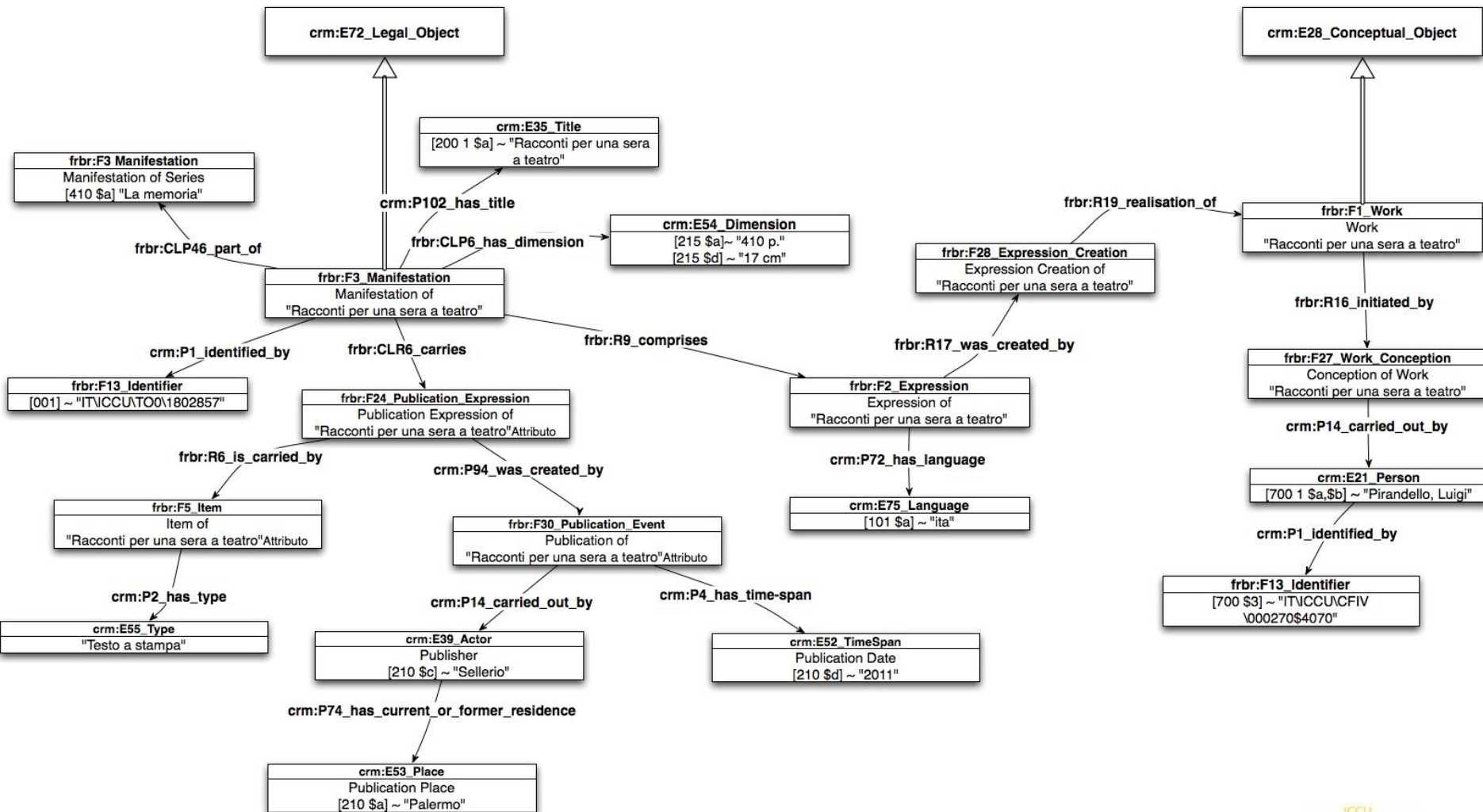
ATHENA PLUS
Access to cultural heritage
networks for Europeana

001 IT\ICCU\TO0\1802857

005 20110418121539.2

010 $a88-389-2536-4

010 $a978-88-389-2536-8

020 aITb2011-882

100 $a20110304d2011 ||||0itac50 ba

101 $aita

102 $aIT

200 1 $aRacconti per una sera a teatro$fLuigi Pirandello$ga cura di Guido Davico Bonino

210 $aPalermo$cSellerio$d2011

215 $a410 p.$d17 cm

410 0$1001IT\ICCU\CFI\0000165$12001 $a<<La >>memoria$v844

676 $a853.914$cNARRATIVA ITALIANA, 1945-1999$v21

700 1$aPirandello$b, Luigi$f <1867-1936>$3IT\ICCU\CFIV\000270$4070

702 1$aDavico Bonino$b, Guido$3IT\ICCU\CFIV\003294

790 1$aBirandillu$b, Luwgi$3IT\ICCU\IEIV\076082$zPirandello, Luigi <1867-1936>

790 1$aPirantelo$b, Louitzi$3IT\ICCU\PUVV\323459$zPirandello, Luigi <1867-1936>

790 1$aPirandelo$b, Luitzi$3IT\ICCU\PUVV\323462$zPirandello, Luigi <1867-1936>

801 3aITbICCU$c20130610

[…]

ATHENA
PLUS
Access to cultural heritage

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

```
<frbr:F3_Manifestation rdf:about="http://www.sbn.it/IT\ICCU\TO0\1802857">


  <crm:P102_has_title>
   <crm:E35_Title>Racconti per una sera a teatro</crm:E35_Title>
  </crm:P102_has_title>


  <frbr:R9_comprises>
   <frbr:F2_Expression rdf:about="http://www.sbn.it/expression/IT\ICCU\TO0\1802857">
    <crm:P72_has_language>
      <crm:E75_Language>ita</crm:E75_Language>
      </crm:P72_has_language>
   </frbr:F2_Expression>
  </frbr:R9_comprises>


[…]


 </frbr:F3_Manifestation>
```

LOD in the framework of the SBN
Patrizia Martini - Achille Felicetti

Thank You!

Contact

Patrizia Martini

patrizia.martini@beniculturali.it

Achille Felicetti

achille.felicetti@gmail.com

www.athenaplus.eu
Twitter | LinkedIn (http://linkd.in/19XwWOz)