MARC into Linked Data: An update on the Bibliographic Framework Initiative

NISO/DCMI Webinar
23 January, 2013

Eric Miller
em@zepheira.com
Goals

• Design of BIBFRAME
• BIBFRAME and Linked Data
• Overview of the BIBFRAME model
• Status and Next Steps
  • and how you can help
BIBFRAME

“The Initiative aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace.”

- Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services
  Library of Congress, November 21, 2012
BIBFRAME

“The Initiative aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace.”

* robust enough to deal with 40 years ago
* flexible enough to deal with 40+ years from now
“The Initiative aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace.”

* libraries, museums, archives, publishers, etc.
* community (users) they serve

Library of Congress, November 21, 2012
“The Initiative aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace.”

- Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services

* not on the web
* but of the web
Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control
Linked Data

"a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

http://en.wikipedia.org/wiki/Linked_Data
Sounds great, what's the catch?
It takes the big step of fundamentally rethinking applications and their integration
Balance

Web Architecture and Standards

Systems Engineering

Business and Change Management

Subject Matter Expertise
BIBFRAME Design Background

• Deconstructing MARC
  • Linked Data “Things not Strings”
• Ground truth of data patterns
• Analysis of existing Linked Data efforts / patterns
• Implementable
• Addresses RDA use cases / scenarios (first among equals)
• Adheres to Web Architecture
• History / experience
MARC as “Things not Strings”

<table>
<thead>
<tr>
<th>001</th>
<th>26132811</th>
</tr>
</thead>
<tbody>
<tr>
<td>008</td>
<td>920608s1992 nyu bf 001 0 eng</td>
</tr>
<tr>
<td>010</td>
<td>$a92021087</td>
</tr>
<tr>
<td>020</td>
<td>$a1555701205</td>
</tr>
<tr>
<td>040</td>
<td>$aDLC$cDLC$dAGL</td>
</tr>
<tr>
<td>049</td>
<td>$aHNKA</td>
</tr>
<tr>
<td>050</td>
<td>00 $aZ678.9$b.C58 1992</td>
</tr>
<tr>
<td>082</td>
<td>00 $a025/.00285$220</td>
</tr>
<tr>
<td>100</td>
<td>1 $aCohn, John M</td>
</tr>
<tr>
<td>245</td>
<td>10 $aPlanning for automation :$ba how-to-do it manual for librarians /$cJohn M. Cohn, Ann L. Kelsey, Keith Michael Fiels</td>
</tr>
<tr>
<td>260</td>
<td>$aNew York :$bNeal Schuman Publishers,$cc1992</td>
</tr>
<tr>
<td>300</td>
<td>$avi, 116 p. ;$c28 cm</td>
</tr>
<tr>
<td>440</td>
<td>0 $aHow-to-do it manuals for libraries ;$uno. 25</td>
</tr>
<tr>
<td>504</td>
<td>$aIncludes bibliographical references and index</td>
</tr>
<tr>
<td>650</td>
<td>0 $aLibraries$xAutomation$xManagement$xHandbooks, manuals, etc</td>
</tr>
<tr>
<td>650</td>
<td>0 $aLibrary planning$xHandbooks, manuals, etc</td>
</tr>
<tr>
<td>700</td>
<td>10 $aKelsey, Ann L</td>
</tr>
<tr>
<td>700</td>
<td>10 $aFiels, Keith Michael</td>
</tr>
</tbody>
</table>
MARC as “Things not Strings”

001 26132811
008 920608s1992 nyu 0 eng
010 $a92021087
020 $a1555701205
040 $aDLC$cDLC$dAGL
049 $aHNKA
050 00 $aZ678.9$b.C58 1992
082 00 $a025/.00285$220
100 1 $aCohn, John M
245 10 $aPlanning for automation :$ba how-to-do it manual for librarians /$cJohn M. Cohn, Ann L. Kelsey, Keith Michael Fiels
260 $aNew York :$bNeal Schuman Publishers,$cc1992
300 $avi, 116 p. ;$c28 cm
440 0 $aHow-to-do it manuals for libraries ;$uno. 25
504 $aIncludes bibliographical references and index
650 0 $aLibraries$xAutomation$xManagement$xHandbooks, manuals, etc
650 0 $aLibrary planning$xHandbooks, manuals, etc
700 10 $aKelsey, Ann L
700 10 $aFiels, Keith Michael
MARC as “Things not Strings”
MARC as “Things not Strings”

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>26132811</td>
</tr>
<tr>
<td>008</td>
<td>920608s1992 nyu</td>
</tr>
<tr>
<td>010</td>
<td>$a92021087</td>
</tr>
<tr>
<td>020</td>
<td>$a1555701205</td>
</tr>
<tr>
<td>040</td>
<td>$aDLC$cDLC$dAGL</td>
</tr>
<tr>
<td>049</td>
<td>$aHNKA</td>
</tr>
<tr>
<td>050</td>
<td>00 $aZ678.9$b.C58 1992</td>
</tr>
<tr>
<td>083</td>
<td>00 $a025/.00285$220</td>
</tr>
<tr>
<td>10 1</td>
<td>$aCohn, John M</td>
</tr>
<tr>
<td>10 10</td>
<td>$aPlanning for automation :$ba how-to-do it manual for librarians /$cJohn M. Cohn, Ann L. Kelsey, Keith Michael Fiels</td>
</tr>
<tr>
<td>260</td>
<td>$aNew York :$bNeal Schuman Publishers,$cc1992</td>
</tr>
<tr>
<td>300</td>
<td>$avi, 116 p. ;$c28 cm</td>
</tr>
<tr>
<td>440 0</td>
<td>$aHow-to-do it manuals for libraries ;$uno. 25</td>
</tr>
<tr>
<td>504</td>
<td>$aIncludes bibliographical references and index</td>
</tr>
<tr>
<td>650 0</td>
<td>$aLibraries,$xAutomation,$xManagement,$xHandbooks, manuals, etc</td>
</tr>
<tr>
<td>650 0</td>
<td>$aLibrary planning,$xHandbooks, manuals, etc</td>
</tr>
<tr>
<td>700 10</td>
<td>$aKelsey, Ann L</td>
</tr>
<tr>
<td>700 10</td>
<td>$aFiels, Keith Michael</td>
</tr>
</tbody>
</table>
MARC as “Things not Strings”
MARC as “Things not Strings”

http://viaf.org/ ...

http://geonames.org/ ...

http://id.loc.gov/ ...

http://...
Ground Truth of MARC data

A hand-picked, 2k set of MARC records from the Library of Congress to illustrate different conditions and kinds of resources that may be found in library databases. MARC fields and conventions used depend upon the form of material being cataloged.

prefix namespace URI
marc http://www.loc.gov/MARC21/slim

Color Codes
- **identifier** - every time the path was found, it contained a different value
- **constant** - every time the path was found, it contained the same value
- **empty** - every time the path was found, it contained no values

Path Info
- **xpath** - name of the element or attribute
- **count** - how many values were found in this path
- **uniques** - how many unique values were found
- **length** - the average character length of the values

```
xpath  |  count | uniques | length
-------|--------|---------|--------
marc:collection | 1 | 1 | 0
marc:record | 2259 | 6543 | 19
  marc:controlfield | 7458 | 4 | 3
    @tag | 7458 | 4 | 3
marc:datafield | 43196 | 8 | 1
  @ind1 | 43196 | 10 | 1
  @ind2 | 43196 | 160 | 3
    @tag | 43196 | 160 | 3
marc:subfield | 97129 | 34545 | 15
    @code | 97129 | 33 | 1
marc:leader | 2259 | 2188 | 24
```
Linked Data in Libraries
Sample BIBFRAME records from the Library of Congress (representative of public library)

1000 Works

<table>
<thead>
<tr>
<th>1. Immanuel Velikovsky - The Truth Behind the Torment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator: Sharon, Ruth Velikovsky</td>
</tr>
<tr>
<td>Subject(s): Velikovsky, Immanuel.</td>
</tr>
<tr>
<td>: Astronomers (✓) / United States</td>
</tr>
<tr>
<td>: Scientists (✓) / United States</td>
</tr>
<tr>
<td>: Astronomers (✓) / Russia (Federation)</td>
</tr>
<tr>
<td>: Scientists (✓) / Russia (Federation)</td>
</tr>
<tr>
<td>: Astronomy (✓) / History.</td>
</tr>
<tr>
<td>: Celestial mechanics.</td>
</tr>
<tr>
<td>Instance(s):</td>
</tr>
<tr>
<td>Immanuel Velikovsky - The Truth Behind the Torment.</td>
</tr>
<tr>
<td>Publisher: London Paradigma Ltd 2010</td>
</tr>
<tr>
<td>ISBN: 9781906833213</td>
</tr>
<tr>
<td>Physical: 200 p. ill. 16 x 23 cm</td>
</tr>
<tr>
<td>Description:</td>
</tr>
</tbody>
</table>

| 2. Country madness : an English country diary of a Singaporean psychiatrist / Ong Yong Lock. |
|                                                                                             |
| Creator: Ong, Yong Lock.                                                                    |
| Note(s): Includes bibliographical references (p. 79-80).                                   |
| Subject(s): Ong, Yong Lock.                                                                 |
|           : Psychiatrists (✓) / Norfolk                                                       |
|           : Singaporeans (✓) / Norfolk                                                        |
|           : Norfolk (England) (✓) / Social life and customs.                                 |
|           : Norfolk (England) (✓)                                                             |
| Instance(s):                                                                                |
|   Country madness : an English country diary of a Singaporean psychiatrist / Ong Yong Lock.|
| Publisher: Singapore : Monsoon, 2010.                                                      |
| ISBN: 9789101854324                                                                         |
| Physical: 80 p. ; ill. (some col.) ; 20 cm.                                                 |
What is Bibframe?

- Bibframe is a linked data model for bibliographic data
- Requirements and Goals:
  - Inherent the robust nature of the MARC standard
  - Be content model agnostic - RDA, DACS, CCO, etc.
  - Handle description and management of traditional (books, serials, etc.) and non-traditional materials (born digital, data sets)
  - Integrate supporting data (Authority, Holdings, Classification)
  - Benefits from all the goodness the Web offers (web triggers, collaboration, data driven, etc.)
**Work (content)**

Resource reflecting the conceptual essence of the cataloged item
**Instance (carrier)**

Resource reflecting a material embodiment of a BIBFRAME Work
Authority

Resources reflecting key authoritative concepts (Person, Place, Org., Topic, etc.). Targets of defined relationships.
Annotation

A means for decorating other BIBFRAME resources with additional information such as Library Holdings, Cover Art, Reviews, etc.
MARC to BIBFRAME

035 (RLIN)MIUG83-B71715
035 (CaOTULAS) 160087264

**Work**
040 MH c MiU c MiU

**Work**
100 1 [Flint, Henry Marty] d 1829-1868.

**Work**
245 00 Life of Stephen A. Douglas; b to which are added his speeches and reports.

**Instance**
260 Philadelphia, b J. E. Potter and co. c [c1863]

**Instance**
300 pp.222, 218. c 20 cm.

**Work**
500 At head of title: Authorized edition.

**Work**
600 10 Douglas, Stephen Arnold, d 1813-1861.
Scenario 1: A new expression for an existing simple work

Link to Scenario 1 RDF Expression and Analysis: ["Scenarios/1"]

Jane Cataloger is assigned to work on a gift collection. Her first selection is a Latvian translation of Kurt Vonnegut's "Bluebeard: a novel." She searches the library database for the original work, and finds:

* Author: Kurt Vonnegut
* Preferred Title for the Work: Bluebeard: a novel.
* Form of Work: Novel

with links to the following expression information:

* Language of Expression: English
* Content Type: Text

and one manifestation:

* Place of Production: New York
* Publisher's Name: Delacorte Press
* Date of Production: 1987
* Extent: 300 pages

Jane begins her description by linking to the existing Work entity. She then creates an expression description:

* Language of Expression: Latvian
* Translator: Arvida Grigulis

She creates an authority record for the translator since none yet existed. She continues by creating a fuller description for the new manifestation, linking to the authority record for the Latvian publisher (what luck, it already existed!).

* Title Proper: [in Latvian]
* Place of Production: Riga
* Publisher: Liesma
* Date of Production: 1997

http://wiki.dublincore.org/index.php/Cataloger_Scenarios
Bluebeard: a novel

Author: Kurt Vonnegut
Summary: An autobiography of Rabo Karabekian, an abstract expressionist artist, who acquired the largest collection of abstract expressionist paintings in private hands.
Rating:  (not yet rated)  Be the first.
Bluebeard (Paperback)
Published September 8th 1998 by Dial Press Trade Paperback
Paperback, 318 pages
more details

Bluebeard (Paperback)
Published October 1st 1988 by Dell Publishing (NYC)
Paperback, 304 pages
more details

Bluebeard (Hardcover)
Published September 2nd 1987 by Delacorte Press
Hardcover, 300 pages
more details

Bluebeard (Paperback)
Published November 1988 by Dell Publishing (NYC)
Paperback, 287 pages
more details

Bluebeard (Paperback)
Published March 16th 1989 by Flamingo
Paperback, 256 pages
more details

Bluebeard (Kindle Edition)
Kindle Edition
more details

Barbablù (Paperback)
Published May 31st 2007 by Feltrinelli
I Narratori, Paperback, 239 pages
more details

Bluebeard (ebook)
Published October 14th 2009 by Dial Press Trade Paperback
ebook, 154 pages
more details
<table>
<thead>
<tr>
<th>Title</th>
<th>Edition</th>
<th>Publisher</th>
<th>Publication Date</th>
<th>Pages</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluebeard</td>
<td>Paperback</td>
<td>Dell Publishing (NYC)</td>
<td>October 1st 1988</td>
<td>304 pages</td>
<td>more details</td>
</tr>
<tr>
<td>Bluebeard</td>
<td>Hardcover</td>
<td>Delacorte Press</td>
<td>September 2nd 1987</td>
<td>300 pages</td>
<td>more details</td>
</tr>
<tr>
<td>Bluebeard</td>
<td>Paperback</td>
<td>Dell Publishing (NYC)</td>
<td>November 1988</td>
<td>287 pages</td>
<td>more details</td>
</tr>
<tr>
<td>Bluebeard</td>
<td>Paperback</td>
<td>Flamingo</td>
<td>March 16th 1989</td>
<td>256 pages</td>
<td>more details</td>
</tr>
<tr>
<td>Bluebeard</td>
<td>Kindle Edition</td>
<td></td>
<td></td>
<td></td>
<td>more details</td>
</tr>
<tr>
<td>Barbablù</td>
<td>Paperback</td>
<td>I Narratori</td>
<td>May 31st 2007</td>
<td>239 pages</td>
<td>more details</td>
</tr>
<tr>
<td>Bluebeard</td>
<td>Ebook</td>
<td>Dial Press Trade Paperback</td>
<td>October 14th 2009</td>
<td>154 pages</td>
<td>more details</td>
</tr>
</tbody>
</table>

Goodreads.com identifies 32 various resources
Broad humor and bitter irony collide in this fictional autobiography of Rabo Karabekian, who, at age seventy-one, wants to be left alone on his Long Island estate with the secret he has locked inside his potato barn. But then a voluptuous young widow badgers Rabo into telling his life story—and Vonnegut in turn tells us the plain, heart-hammering truth about man’s careless fancy to create or destroy what he loves.

"id": "http://bibfra.me/work/u2.t1",
"title": "Barbablù",
"type": "Book",
"author": "http://bibfra.me/auth/person/kurt_vonnegut",
"translator": "http://bibfra.me/auth/person/pier_francesco_paolini",
"subject": ["http://bibfra.me/auth/subject/foo","http://bibfra.me/auth/subject/bar","http://bibfra.me/auth/subject/baz"],
"language": "Italian",
"hasInstance": "http://bibfra.me/instance/u2.t1-1",
"isTranslationOf": "http://bibfra.me/work/u2"

"id": "http://bibfra.me/instance/u2.t1-1",
"type": "PaperbackBook",
"volumes": "1",
"pages": "239",
"date": "2007-05-31",
"publisher": "http://bibfra.me/auth/org/feltrinelli",
"ISBN": "880701727X",
"image": "http://d.gr-assets.com/books/1327279411l/9700830.jpg"
Recombinant Data
• Resource Description Framework
• common model for representing web data
RDF

- Resource Description Framework
- common model for representing web data
• Resource Description Framework

• common model for representing web data

http://z.com  dc:creator  http://me.com
RDF

P = 8.0 mm
= 5/6 \times H
= 2.5 \times h

5.0 mm

1.7 mm

3.0 mm

H = 9.6 mm
= 1.2 \times P
= 3 \times h

h = 3.2 mm
= 0.4 \times P
= 1/3 \times H

2 \times P - 0.2 mm
= 15.8 mm

P - 0.2 mm
= 7.8 mm
Community Profiles
Community Profiles
Community Profiles
Back to Bluebeard (in an Open World)

Book isTranslationOf Book

hasInstance

Hardcover (1987)

hasInstance

Paperback (2007)
Incoming Actions for Bluebeard: a novel

- 2007-02-03: New translation work created Barbablù
  would you like to update your resource to connect to it?
- 2007-02-03: New Collection created Bluebeard Collection
  would you like to update your resource to connect to it?
Incoming Actions for *Bluebeard: a novel*

- 2007-02-03: New translation work created *Barbablù*
  would you like to update your resource to connect to it?

- 2007-02-03: New Collection created *Bluebeard Collection*
  would you like to update your resource to connect to it?
Initial Phase

- Completed Landscape Assessment
- Initial draft of Linked Data model *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*
- Prototype of translation services and Interfaces
  - [https://github.com/lcnetdev/marc2bibframe](https://github.com/lcnetdev/marc2bibframe)
- Vocabulary mappings to MARC and RDA
- Project roadmap and deployment strategic report
- Convened Early Experimenters to help refine this work
Early Experimenters

- British Library
- Deutsche Nationalbibliothek
- George Washington University
- National Library of Medicine
- OCLC
- Princeton University
- Library of Congress
Early Experimenters Process

- Face to Face meeting in October at the Library of Congress to introduce the proposed BIBRAME model and preview their data in the model
- Experimentation for a month at their institutions
- Second Face to Face meeting in December to share findings, comments and questions
- Iterative refinement of model, vocabulary and services based on real world data and experimentation
Next phase

- Extended Experimentation with more institutions and data types
- Create more detailed documentation
- Complete Vocabulary Mapping to address all MARC and RDA fields
- Serialization and Protocols
- Create Use Cases for further validation
- Named Profiles / constraints
- Prototype lightweight, open source cataloging interfaces
- Generate information central for all things BIBFRAME
BIBFRAME: Proposed Roadmap

Short-Term Goal:
Refine BIBFRAME Model and Prepare for Transition

Medium-Term Goal:
Community Education and Transition

Long-Term Goal:
Adoption

BIBFRAME Model Development
- Landscape Assessment
- Early Experiments
- BIBFRAME Model Refinement & Application Profiles
- Pipeline Development
- Model Development
- Annotation Demonstration and Stress Testing

Community/Market Adoption
- Education & Outreach Materials
- Stakeholder Transition Strategy, Education and Transition
- Library Vendor and Search Strategy
- Library Vendor and Search Adoption
- Use Cases and Validation
- End User Facing Tool Development
- Value Proposition for Transition
- Authority Services
- BIBFRAME Vocabulary Governance Model

Completed Tasks

Next Steps for you

• Continue to learn about Linked Data (inside and outside of Library community)

• Join the discussion
  • bibframe@listserv.loc.gov

• Keep a watch out for updates to documentation, examples, vocabulary, services, prototypes, etc.
  • http://www.loc.gov/marc/transition/

• Start to engage vendors
  • “what is your Linked Data strategy”?
Thank You

Comments or Questions

Eric Miller
em@zepheira.com