Library Linked Open Data

Why, What, and Where...

Jackie Shieh
Descriptive Data Management
ShiehJ [at] si.edu

Smithsonian Libraries
Outline

• Why Linked data
• What is Resource Description Framework (RDF)
• What is a Uniform Resource Identifier (URI)
• Where to place URIs in MARC Bibliographic Records
• What skills will be needed
Linked data describes a method of publishing structured data so that it can be interlinked and become more useful.

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

- A roadmap to the Semantic Web by Tim Berners-Lee
Data graph principles

Use URIs for names for THINGs
Use HTTP URIs for lookup
Use standards (e.g. RDF)
Include links to other URIs

https://www.w3.org/DesignIssues/LinkedData.html
Topic

• What is Resource Description Framework (RDF)
• What is Uniform Resource Identifiers (URIs)
• Where to place URIs in MARC Bibliographic Records
• What skills will be needed
A RDF statement

A RDF statement consists of three parts:
- **Subject**: Explicit
- **Predicate**: controlled from any vocabulary
- **Object**: literal value or another resource
Statement examples

Michael
subject
knows
predicate
Michael
subject
is-a
predicate

David
object
boy
object
RDF Triples and URI

Subject is usually a URI

Predicate is always a URI

Object can be a URI

Object can also be a literal (string)
Topic

• What is Uniform Resource Identifiers (URIs)
• Where to place URIs in MARC Bibliographic Records
• What skills will be needed
A Uniform Resource Identifier (URI) is a string of characters that unambiguously identifies a particular resource. To guarantee uniformity, all URIs follow a predefined set of syntax rules,[1] but also maintain extensibility through a separately defined hierarchical naming scheme (e.g. http:// ).

<table>
<thead>
<tr>
<th>Scheme/Protocol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftp</td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td>http/https</td>
<td>Hyper text transfer protocol (Secure)</td>
</tr>
<tr>
<td>mailto</td>
<td>Electronic mail address</td>
</tr>
<tr>
<td>ssh</td>
<td>Secure shell</td>
</tr>
<tr>
<td>tel</td>
<td>telephone</td>
</tr>
<tr>
<td>urn</td>
<td>Uniform Resource Names</td>
</tr>
</tbody>
</table>
Catalog record example

Pride and prejudice / Jane Austen.

Record info:
Main Author: Austen, Jane, 1775-1817.
Published/Created: 2007 - Ann Arbor, MI : Borders Classics,
Language: English
Subjects (LCSH): Courtship -- Fiction.
Entitification of library data

Subject: Jane Austen

Dates: 1775-1817

Type: Person

Born in: Hampshire, England

Type: County

authored

 Predicate

Object: Pride and Prejudice

Type: Book

Topic: Courtship

isPartOf

Type: United Kingdom

isIn

Type: Great Britain

Type: Hampshire, England

isPartOf

Type: United Kingdom

Type: Country

isPartOf

Type: United Kingdom

Type: Country
HTTP URIs in entity

- **Subject Identifier**: http://viaf.org/viaf/102333412
- **Predicate**: http://www.wikidata.org/entity/Q21694653
- **Object Identifier/Value**: http://id.loc.gov/authorities/names/n2002041181

**Birth and Death Dates**
- 1775
- 1817

**Locations and Types**
- https://www.wikidata.org/entity/Q215627
- https://www.wikidata.org/entity/Q6581072
- https://www.wikidata.org/entity/Q21694653
- https://www.wikidata.org/entity/Q145
- https://www.wikidata.org/entity/Q171809
- http://viaf.org/viaf/127756949
- http://viaf.org/viaf/127756949

**Identification and Relationships**
- http://purl.org/dc/terms/creator
- http://schema.org/author
- http://id.loc.gov/authorities/subjects/sh85033596
- http://id.loc.gov/authorities/names/n2002041181
- http://schema.org/Book
- http://viaf.org/viaf/127756949
- http://id.loc.gov/authorities/subjects/sh85033596
- isPartOf
- isIn
- Topic

**Additional Information**
- http://www.wikidata.org/entity/Q21694653
- http://www.wikidata.org/entity/Q145
Not all URIs are equal

2XX: Request successfully received, understood and accepted

3XX: Request will complete after further action

301 (Redirect)
303 (See other)

https://www.w3.org/TR/cooluris/
<table>
<thead>
<tr>
<th>URI known as URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dereferenceable URI</td>
</tr>
</tbody>
</table>

**Canonical Concept URI**

http://vafu.redlink.io/
Where to locate RDF URIs

- **SPARQL** (**S**imple **P**rotocol **A**nd **RDF** **Q**uery **L**anguage)
- **Endpoints** (vocabularies)
  - Library of Congress: id.loc.gov
  - OCLC FAST: fast.oclc.org
  - Wikidata: query.wikidata.org
  - Etc.

*See PCC URIs in MARC Task Group’s document: [Formulating and obtaining URIs](http://bit.ly/SLA2019Shieh_l lod)*
Data evolution

Tabular Database
- Data stored in Tables
- Tables are related by Key Fields

Tree Data (XML/JSON)

Data graph (RDF)

jerry
```
Jersey
```
tom

Mouse
Cat

subClassOf

Rodentia

subClassOf

Mammalia

ce.utexas.edu

stackoverflow.com

Viewpointusa.com
Topic

• Where to place URIs in MARC Bibliographic Records

• What skills will be needed
MARC bibliographic data

Where should URIs/IRIs for RDF objects go?

- $u$ URL ➔ Uniform Resource Identifier
- $0$ Authority record control number or standard number
- $1$ (uri) http://www.wikidata.org/entity/Q36322
$0 in Connexion

100 1#$aBach, Johann Sebastian.$a4aut$0(DE-101c)310008891

http://d-nb.info/gnd/310008891

600 1 7 Austen, Jane,$d1775–1817.$d fast $0 (OCoLC)fst00032929
655 7 Personal correspondence.$d fast $0 (OCoLC)fst01919948
Identifier (ID) – URL – IRI (URI) for Carrier type

**Toolkit profiles**

Elements specified by:

**ID (local):**
ala-0dd63c0d-066f-3ddc-885a-ff83c25cf752

**URL (global):**
https://beta.rdatoolkit.org/RDA.Web/Content?externalId=en-US_ala-0dd63c0d-066f-3ddc-885a-ff83c25cf752

**IRI (global):**
http://rdaregistry.info/Elements/m/P30001
Topic

• What skills will be needed
SIL FAST headings project

MARC Bib (selective set) → FAST Headings? → Yes → RDF URI in $0/$1? → Yes → RDF URIs in BIB → NO → NO → Conform FAST Heading Identifier to RDF URI → Add FAST Headings → NO → RDF URIs in BIB
Bib data enhancements

• Added FAST identifiers (then converted to URIs)
• Updated MARC encoding standards and data validation
• RDA content standards
• Added URIs to AAPs and alternates (1xx, 6xx, 7xx)
• Replaced/added relationship URI ($4) to relator code
• Removed ISBD punctuation (trailing and between subfields)
Authority data enhancements: challenges

• System challenges in loading multiple authority files

• Local practice: authority control at point of cataloging
  • Single item authority workflow
  • Batch process authority workflow

• Post-load reconciliation process (e.g. VIAF)
Identity management: opportunities

• Semantic relations: e.g., “sameAs”, “relatedMaterials”, “closeMatch”
• Coordinates and GeoNames
• Creating new identities using Wikidata (via local instance of Mediawiki/Wikibase)
Staff impacts: breaking old habits

• Cataloging silos
  • Multiple cataloging sites with diverging policies
  • Communication struggles

• Training inconsistencies

• Reinventing workflow
  • Development for all staff levels
  • Connecting new technologies to daily tasks
For want of a nail

Windows 10
- Cataloging
  - Connexion
  - Horizon
- Microsoft Office 365
  (Excel + ASAP Utilities, Powerpoint screen recording)
- cURL (client URL)

Open Source
- MarcEdit
- TextPad
- OpenRefine
- Draw.io

Linux
- MARC::Record Perl package
- Yaz toolkit (Index Data)
- Perl and shell scripts
- cURL (client URL)
Staff impacts: learning new skills

- Regular Expression concept
- MarcEdit
  - MARCJoin
  - Validate MARC records
  - Find/replace
  - Manage/apply tasks
  - Export tab delimited records
  - Export/import for/from OpenRefine
  - Remove blank lines
  - RDA helper
  - Build linked records
  - Compile file into MARC
- OpenRefine
  - Clustering
  - Using “blank down” function to remove duplicates
- TextPad
  - Free evaluation copy
  - Has block select functionality
  - Sort and dedupe
Some References

• What is RDF?
  www.w3schools.com/XML/xml_rdf.asp

• ALCTS Webinars: *From MARC to BIBFRAME*
  ala.org/alcts/confevents/upcoming/webinar/MARCtoBIBFRAME

• PCC URIs in MARC Task Group
  loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html

• URI FAQs
  loc.gov/aba/pcc/bibframe/TaskGroups/URI%20FAQs.pdf

• Formulating and Obtaining URIs
  loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf

• MARC subfield $0$
  loc.gov/marc/bibliographic/ecbdcntf.html
Thank you!

Jackie Shieh
Descriptive Data Management
ShiehJ [at] si.edu

Smithsonian Libraries