Demystifying Linked Data: A Practical Guide For Librarians and Staff

Eric Hanson
Electronic Resources Librarian
Acquisitions & Discovery Department
North Carolina State University Libraries
This Presentation

• Brief overview of RDF and linked data concepts

• Primary focus will be on the practical steps librarians and staff can take to publish linked data
Overview

- RDF (Resource Description Framework) is the data model used in linked data
- RDF triples consist of Subject — Predicate — Object

**Example**

The Lancet — Has Publisher — Elsevier
(Subject) (Predicate) (Object)
Overview

• Subjects, predicates, and objects in RDF triples are generally represented with URIs (Uniform Resource Identifiers)

Example

The Lancet — Publisher — Elsevier
(Subject) (Predicate) (Object)
Overview

• Linked data is created using RDF vocabularies, which contain classes and properties

• Classes are types of things (Person, Document, Organization)

• Properties describe things (title, date, creator)

• By using the same RDF data model, it is easy to combine information from multiple data sets
Linked Data Creation Process

• Model
• Clean Up
• Enhance
• Convert
• Publish

This is modified and simplified version of the process described in:
NCSU Organization Name Linked Data

- [http://www.lib.ncsu.edu/ld/onld/](http://www.lib.ncsu.edu/ld/onld/)
- Contains over 2,100 organizations
- Based on the NCSU Organization Name Authority, a tool used to manage the variant forms of name for publishers, providers, and vendors in our E-Matrix ERMS

Diagram:
- **Royal Society of Chemistry**
  - Authorized
- Royal Society of Chemistry (RSC)
  - Variant
- Royal Society of Chemistry Publishing
  - Variant
- The Royal Society of Chemistry
  - Variant
Model

• Find properties and classes from RDF vocabularies to describe your data set

• General vocabularies

  Dublin Core  http://dublincore.org/documents/dcmi-terms/
  RDF Schema  http://www.w3.org/TR/rdf-schema/
  SKOS        http://www.w3.org/2009/08/skos-reference/skos.html
  OWL         http://www.w3.org/TR/owl-ref/
  FOAF        http://xmlns.com/foaf/spec/
Model

• Search for more specific vocabulary terms when necessary

Open Metadata Registry [http://metadataregistry.org/](http://metadataregistry.org/)
Linked Open Vocabularies [http://lov.okfn.org/dataset/lov/](http://lov.okfn.org/dataset/lov/)
Model

- Modeling example from NCSU Organization Name Linked Data

**Authorized Name**
Royal Society of Chemistry

**Variant Names**
Royal Society of Chemistry (RSC)
Royal Society of Chemistry Publishing
The Royal Society of Chemistry

---

**skos:prefLabel**
preferred label

**skos:altLabel**
alternative label
Clean Up

- Make data as clean as possible before trying to convert to RDF
- Types of errors vary with each data set: duplicate entries, inconsistent formatting, diacritics (Ö, ó, 社団, Č)
- Make sure data is in a structured format such as XML, can be done with tools such as MS Access, MarcEdit
- Do your best before conversion, but data clean up will likely happen throughout the conversion process as errors become visible in new data formats
Enhance

• Links to other data sets create context and allow users to easily combine information from multiple data sets

• Tim Berners-Lee’s 5 Star Open Data

  http://www.w3.org/DesignIssues/LinkedData.html

Oxford University Press

VIAF  DBpedia  ISNI

LCNAF  Freebase
http://id.loc.gov/authorities/names/n80126136  http://rdf.freebase.com/ns/m/0czzl
Enhance

- Batch processing: Using **OpenRefine** with the **RDF extension** to search against other data sets
Enhance

- Search linked data sets manually for equivalent URIs
Convert

• Many tools for converting data to RDF: XSLT, OpenRefine, Python, JavaScript, Java, etc.

• We used XSLT to transform our data into various RDF formats

• XSLT is easy to learn for those with HTML and XML experience
Convert

• Don’t know XSLT or other scripting languages? Learn for free!
  
  W3Schools http://www.w3schools.com/
  Codeacademy http://www.codecademy.com/

• Troubleshooting specific problems?
  
  Stack Overflow http://stackoverflow.com/
  Users helping each fix coding problems
Convert

• Choose what RDF serializations to use for publishing

• We looked at linked data sets from OCLC, the Library of Congress, and Dbpedia to see what forms of RDF were used

• The following are just snippets, consult sample files and serialization specifications for more details
Convert

N-Triples  http://www.w3.org/TR/n-triples/

Convert RDF-XML

http://www.w3.org/TR/rdf-syntax-grammar/

.rdf:Description rdf:about="http://www.lib.ncsu.edu/ld/onld/00000008">
    <skos:prefLabel>Oxford University Press</skos:prefLabel>
    <foaf:homepage rdf:resource="http://global.oup.com/"/>
</rdf:Description>
N3/Turtle  http://www.w3.org/TeamSubmission/n3/ ,
http://www.w3.org/TR/turtle/

@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#>

<http://www.lib.ncsu.edu/ld/onld/00000008>
    skos:prefLabel "Oxford University Press" ;
Convert

JSON-LD  http://json-ld.org/

{
    "@context":
    {
        "owl": "http://www.w3.org/2002/07/owl#",
        "foaf": "http://xmlns.com/foaf/0.1/",
        "owl:sameAs": {
            "@type": "@id"},
        "skos": "http://www.w3.org/2004/02/skos/core#"},

    "@graph": [
        {
            "@id": "http://www.lib.ncsu.edu/ld/onld/00000008",
            "skos:prefLabel": "Oxford University Press",
            "foaf:homepage": "http://global.oup.com/"
        }
    ]}
Convert

HTML with RDFa - Browser View

Oxford University Press

Authorized Name

skos:prefLabel Oxford University Press

Website

foaf:homepage http://global.oup.com/
Convert HTML with RDFa - HTML View

```html
<h2 style="font-weight:bold;">Oxford University Press</h2>
<br/>
<table>
  <tr>
    <td><a href="http://www.w3.org/2004/02/skos/core#prefLabel">skos:prefLabel</a></td>
    <td><span property="skos:prefLabel" style="font-weight:bold;">Oxford University Press</span></td>
  </tr>
</table>
<br/>
<h3>Website</h3>
<br/>
<table>
  <tr>
    <td><a href="http://xmlns.com/foaf/0.1/homepage">foaf:homepage</a></td>
  </tr>
</table>
```
Publish

• Make your data set available on the web with an appropriate open data license (CC0, ODC-BY, ODbL)

• Announce to listservs and colleagues

• Register your data sets with linked data registries such as http://www.datahub.io

• Provide contact information so users can report errors

• Create plan for updating and maintaining the data set
Publish

• Show your work!

• Linked data is about sharing information so make your process and tools available for others to use as a model

• We made our XSLT files and some sample XML data available for download on the project website
Going forward

• Try a simple project as a learning opportunity

• The skills learned and tools built for your first project will make the second project much easier

• Librarians have a chance in these early stages of linked data to help establish best practices

• Librarians have been working with authority control for decades and we should share our expertise as linked data moves forward
Going forward at NCSU

- NCSU Libraries is collaborating with developers of the Global Open Knowledgebase (GOKb)

- Plans for linked data release involving data about serial and e-resource titles, packages, platforms, and organizations
Questions?

emhanso2@ncsu.edu