8-9-2018

Linked data in the library & OpenRefine

Sai Deng
University of Central Florida, sai.deng@ucf.edu

Find similar works at: https://stars.library.ucf.edu/ucfscholar

University of Central Florida Libraries http://library.ucf.edu

Original Citation
Linked Data in the Library & OpenRefine

Sai Deng, University of Central Florida Libraries
Linked Data

• Questions to think about:
  • Why does linked data matter? How can libraries and librarians embrace linked data?

• Linked Data in the Library and the Web
  • Cases
  • Standards and Vocabularies

• Linked Data Design Principles, Building Blocks & Creation

• Current Digital Repositories and Linked Data

• Data Editing Tool: OpenRefine (Features, Examples)

• Possible Ways to Get Involved
Linked Data

- Why does linked data matter?
  - Change of the web:
    - Before the web (document silos)
    - Web 1.0 (web of documents)
    - Web 2.0 (social web; application silos)
    - Web 3.0 (semantic web)
    - Web 4.0 (intelligence web)
  - Change from web of “documents” to web of “data”/ “Things” (deconstruct webpages; web semantics: person, place, event, creativeWork...);
  - Allow computers to understand and process data on the web.

- What is linked data?
  - It is recommended best practice for exposing, sharing, and connecting data, information, and knowledge (structured data) on the web;
  - Key technologies: URIs, HTTP, RDF. — Wikipedia
  - Linked data vs. Semantic Web: Pieces vs. whole

- Resource Description Framework (RDF): Represent information in triples
  - **Subject**: a resource, identified with a URI
  - **Predicate**: specification of the relationship
  - **Object**: a resource the subject is related to
Linked Data Examples

Mansfield Park [abridged]

Jane Austen

Is written by
schema:creator

Is an example of
schema:exampleOfWork

Is written by
schema:creator

Is contributed by
schema:contributor

Is about
schema:about

Young women

England

Country homes

1800-1899

Anna Bentinck

Mansfield Park

Northanger Abbey

The novels of Jane Austen : Pride and prejudice

Persuasion

JANE AUSTEN : the complete novels.

bgn:AudioBook

Has format
schema:bookFormat

*Based on information from worldcat.org. LOD links from this site may change.
Linked Data Examples

• In RDF/XML, e.g.,
  <rdf:Description rdf:about="http://www.worldcat.org/oclc/930255038">
   <schema:name xml:lang="en">Mansfield Park</schema:name> ...
   <schema:about rdf:resource="http://experiment.worldcat.org/entity/work/data/1120181956#Topic/young_women"/>
   <schema:bookFormat rdf:resource="http://bibliograph.net/AudioBook"/>...
  </rdf:Description>...

• In JSON-LD, e.g.,
  
  
  
  {
    "@id" : "http://experiment.worldcat.org/entity/work/data/1120181956#Topic/young_women",
    "@type" : "schema:Intangible",
    "name" : {
      "@language" : "en",
      "@value" : "Young women"
    }
  },...

• In N-triples, e.g.,
  <http://www.worldcat.org/oclc/930255038>
  <http://schema.org/about>
  <http://experiment.worldcat.org/entity/work/data/1120181956#Topic/young_women>...
Linked Data and the Library Community

• What have libraries been doing with linked data in recent years?
  • Preparing for linked data:
    • Add linked open data (LOD) links to bibliographic records (e.g., $0 in MARC records);
    • Control terms and add LOD links to metadata in digital repositories.

• Convert bibliographic and authority records into linked data set:
  • WorldCat Linked Data, e.g., Works;
    Explore: http://experiment.worldcat.org/entity/work/data/1080130963
  • Linked Data for Production (LD4P): prepared data for SHARE-VDE

• Collaborate to develop tools, discovery systems and collections;

• Promote ORCID, ISNI and Identity Management;

• Use open data and linked open data to support research;

• Resources
Linked Data Vocabularies

• Names, Subjects, Thesauri and Value Vocabularies
  • LC Linked Data Service: Authorities and Vocabularies https://id.loc.gov/
  • FAST: Faceted Application of Subject Terminology http://id.worldcat.org/fast
  • GeoNames http://www.geonames.org/ontology/documentation.html
  • DOI (as linked data) http://dx.doi.org
  • Virtual International Authority File (VIAF) https://viaf.org/
  • Worldcat Identities www.worldcat.org/identities
  • International Standard Name Identifier (ISNI) http://www.isni.org/
  • Open Researcher and Contributor ID (ORCID) https://orcid.org/

• Metadata Element Sets and RDF Vocabularies
  • DCMI Metadata Terms http://dublincore.org/documents/dcmi-terms/
  • Creative Commons Rights Expression Language (CC REL)
    https://wiki.creativecommons.org/wiki/CC_REL
  • Friend-of-a-Friend http://www.foaf-project.org/
  • Metadata Authority Description Schema in RDF http://www.loc.gov/standards/mads/rdf/
  • Schema.org Vocabulary http://schema.org/docs/gs.html#schemaorg
  • Simple Knowledge Organization Systems (SKOS) https://www.w3.org/2004/02/skos/

• The Linked Open Data Cloud http://linkeddata.org/
Linked Data Implementations & Resources

• Libraries, Archives and Museums Community
  • Europeana
  • DPLA
  • LIBRIS
  • British National Bibliography
  • Open Library
  • BIBFRAME
    • SHARE-VDE (Virtual Discovery Environment)
  • OCLC Linked Data Wikibase
  • Library.Link Network

• The Web and the Library Community
  • DBpedia
  • Google Knowledge Graph
  • Facebook’s Open Graph
  • Schema.org
    • Bibliographic Extension
    • Extend Schema.org for archives

• Other Linked Data Resources & Implementations
  • DATA.GOV
  • Data Linkage Infrastructure - Census Bureau
  • LinkedGeoData.org

• Small Scale Linked Data Projects
  • University of North Texas Libraries: UNT Names
  • NCSU Libraries: Organization Name Linked Data
Linked Data Implementations

- Google Knowledge Graph

Quotes

There is no charm equal to tenderness of heart.

It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.

Ah! There is nothing like staying at home, for real comfort.

Books

Pride and Prejudice 1813
Emma 1817
Persuasion 1817
Sense and Sensibility 1811
Northanger Abbey 1817

People also search for

Emma Thompson
Charlotte Brontë
Charles Dickens
Emily Brontë
Virginia Woolf
Library Standards Created for Linked Data Environment

• IFLA Library Reference (LRM) Model ([https://www.ifla.org/publications/node/11412](https://www.ifla.org/publications/node/11412))
  • Consolidation of the separately developed IFLA conceptual models: FRBR, FRAD, FRSAD

• Resource Description and Access (RDA)
  • “RDA is a package of data elements, guidelines, and instructions for creating library and cultural heritage resource metadata that are well-formed according to international models for user-focussed linked data applications.”
  • RDA Registry ([http://www.rdaregistry.info/](http://www.rdaregistry.info/))

• Built upon the Standards
  • Make bibliographic information discoverable on the web
  • Linked Open Data Bibliographic System: BIBFRAME
  • Prototype: SHARE-VDE (Virtual Discovery Environment)
IFLA LRM entities

IFLA Library Reference (LRM) Model

- Res
- Work
- Expression
- Manifestation
- Item
- Agent
- Collective Agent
- Person
- Nomen
- Place
- Time-span

* This graph is from: Dunsire, Gordon. Telling tails: metadata standards and the digital humanities.
RDF Graph for RDA Class Hierarchies

RDF graph of the class hierarchy of RDA Entity

RDF graph of the class hierarchy of RDA Agent

Source: http://www.rdaregistry.info/rgAbout/rdaont/ontHierarchies.html
IFLA LRM and RDA entities

Any RDA Thing: Covers all other types of entity

RDA Entity

has appellation

Nomen

Place

Time-span

Agent

is sub-class of

is created by

W

is modified by

I

Res

is sub-class of

is associated with

is sub-class of

Collective Agent

P*

RDA refines LRM relationships as element sub-types (RDF sub-properties)

EMC

* This graph is from Dunsire, Gordon. LRM-RDA. 2017.
Linked Data Implementations
An Instance (in Share VDE)

Book (Instance) in Share VDE

SHARE Virtual Discovery Environment (VDE)
Person (Entity)

Jane Austen (Entity) in Share VDE
http://www.share-vde.org/sharevde/searchNames?n_cluster_id=2568128

- ISNI
- Wikidata
- LC Name Authority File
- WorldCat Identities
- Data.bnf.fr
- VIAF
Work (Entity) in Share VDE
http://share-vde.org/sharevde/searchTitles?t_cluster_id=1092450
Linked Data Design Principles & Five Star-schema of Linked Open Data

• The Four Design Principles of Linked Data (Tim Berners-Lee. 2006)
  • Use Uniform Resource Identifiers (URIs) as names for things.
  • Use HTTP URIs so people can look up those names.
  • When someone looks up a URI, provide useful information using the standards (RDF, SPARQL).
  • Include links to other URIs so that they can discover more things.
  • Source: https://www.w3.org/DesignIssues/LinkedData.html

• Five star-schema of Linked Open Data (Tim Berners-Lee)
  • URIs for naming things, RDF for describing things, SPARQL for querying data; link to other URIs to find more things.
Linked Data Building Blocks

• The LOD Data Model: RDF triples

• Content Rules: relies on external vocabularies and ontologies

• Metadata schema
  • RDF Schema
  • Vocabulary Building Blocks: SKOS, OWL
  • Open Archives Initiative Object Reuse and Exchange (OAI-ORE)

• Serializations of LOD
  • RDF Notation-3/N3, Turtle, N-Triples, JSON-LD, RDFa

• Exchanging LOD
  • SPARQL

• Mitchell, Erik T. Building Blocks of Linked Open Data in Libraries. *Library Technology Reports* (vol. 49, no. 5)
Linked Data Building Blocks

- Take DBpedia as an example:
  - The LOD Data Model: RDF triples
  - Content Rules
    - DBpediaOntology Classes (http://mappings.DBpedia.org/server/ontology/classes/)
  - Metadata schema
    - RDF Schema, XML Schema
    - Vocabulary Building Blocks: FOAF, DC, OWL, SKOS
  - Serializations of LOD
    - Turtle, N-Triples, Trix, RDF/JSON
  - Exchanging LOD
    - SPARQL

- A Dbpeida page: (rdf)About: (Person) Jane Austen (http://DBpedia.org/page/Jane_Austen)
  - dbo:abstract, dbo:birthDate (dbr:Hampshire), dbo:deathDate...
  - dct:subject (e.g., dbc:English_women_novelists)
  - rdf:type (e.g., owl:Thing, foaf:Person, dbo:Agent, dbo:Writer)
  - rdfs:comment
  - owl:sameAs wikidata:Jane Austen
    http://viaf.org/viaf/102333412
  - foaf:gender (female (en) )
  - is dbo:author of (e.g., dbr:Emma (novel))
  - is dbo:influenced of (e.g, dbr:George Crabbe)

Vocabularies from:
DBpedia, dcterms, rdf, owl, foaf...
**Linked Data: SPARQL**

- **SPARQL Explorer** ([http://DBpedia.org/snorql/](http://DBpedia.org/snorql/))
- A very simple example: People who were born in Hampshire before 1900

```sparql
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX: <http://DBpedia.org/resource/>
PREFIX DBpedia2: <http://DBpedia.org/property/>
PREFIX DBpedia: <http://DBpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

WHERE { ?person dbo:birthPlace :Hampshire .
  ?person foaf:name ?name .
  FILTER (?birth < "1900-01-01"^^xsd:date) . }
ORDER BY ?name
```

- Returned results, e.g.,

<table>
<thead>
<tr>
<th>name</th>
<th>birth</th>
<th>death</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Adolphus James Sparrow&quot;@en</td>
<td>&quot;1869-05-10&quot;^^xsd:date</td>
<td>&quot;1936-09-06&quot;^^xsd:date</td>
<td>&quot;Adolphus James Sparrow (10 May 1869 – 6 September 1936) was an English cricketer who represented Hampshire in one first-class match in 1904 against Leicestershire. In his only first-class innings Sparrow made 1 run before being dismissed by John King.&quot;@en</td>
</tr>
<tr>
<td>&quot;Jane Austen&quot;@en</td>
<td>&quot;1775-12-16&quot;^^xsd:date</td>
<td>&quot;1817-07-18&quot;^^xsd:date</td>
<td>&quot;Jane Austen (/ˈdʒeɪn ˈɒstɪn/; 16 December 1775 – 18 July 1817) was an English novelist known primarily for her six major novels which interpret, critique and comment upon the British landed gentry at the end of the 18th century...&quot;@en</td>
</tr>
</tbody>
</table>
Linked Data Creation

• Providing Linked Data: creating, interlinking, publishing

• Creating Linked Data
  • Extract data from spreadsheet, databases, text;
    • Tools: OpenRefine, W3C R2RML, DBpediaSpotlight
  • Using URIs to name entities;
  • Selecting vocabularies
    • Linked Open Vocabularies (LOV) (https://lov.linkeddata.es/dataset/lov/)
    • Expressing data using RDF data model;

• Interlinking Linked Data
  • Links at instance level (rdfs:seeAlso, owl:sameAs) or schema level (SKOS:closeMatch, exactMatch, relatedMatch...)
  • Tool: Linked Data Integration Framework - Silk
A Test: Annotate using **DBpediaSpotlight**

**Log**

**Place of Birth**  Dadeville (Ala.)
**Home State**  Alabama
**War or Conflict**  World War II
**Location of Service**  Guam; Palau; China; Okinawa Island (Japan); Korea; Quantico (Va.); Camp Pendleton (Calif.); Monterey (Calif.)
**Highest Rank**  Lieutenant Colonel; LtCol
**Note**  Went to Naval Post Graduate School and Headquarters Marine Corps.
**City of Birth**  Dadeville
**Country of Birth**  United States
**Location of Interview**  Winter Park, Florida

*Need review. Can use other LOD vocabularies. Works with **Name-Entity Extraction** in OpenRefine.*
### Linked Data Creation

**A Test: Create RDF from Spreadsheet with OpenRefine**

<table>
<thead>
<tr>
<th>id</th>
<th>author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jane Austen</td>
</tr>
<tr>
<td>2</td>
<td>Ernest Hemingway</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>authorid</th>
<th>bookid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

**OpenRefine RDF Extension:**

Edit RDF Skeleton.

**Export:** RDF as Turtle.

```turtle
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://example.org/author/1> a foaf:Person ;
foaf:name "Jane Austen" ;
owl:sameAs <http://dbpedia.org/page/Jane_Austen> .

<http://example.org/book/1> a <http://example.org/Book> ;
dc:title "Northanger_Abbey" ;
owl:sameAs <http://dbpedia.org/resource/Northanger_Abbey> .

<http://example.org/author/2> a foaf:Person ;
foaf:name "Ernest Hemingway" ;
owl:sameAs <http://dbpedia.org/page/Ernest_Hemingway> .

<http://example.org/book/4> a <http://example.org/Book> ;
dc:title "The_Old_Man_and_the_Sea" ;
owl:sameAs <http://dbpedia.org/resource/The_Old_Man_and_the_Sea> .
```

Refer to: From Excel file to RDF with links to DBpedia and Europeana ([https://www.youtube.com/watch?v=XdpzmGxA33U](https://www.youtube.com/watch?v=XdpzmGxA33U))
Linked Data Publishing

• Publishing Linked Data
  • Dataset metadata creation;
    • Describing LD with VoID Vocabulary https://www.w3.org/TR/void/

• Publish dataset via RDFa, SPARQL endpoints;
  • Dereferencing HTTP URIs (e.g., DBpedia)
  • RDFa: embed RDF within HTML (e.g., <div class="artistheader" about=RDF link...>...</div>);
  • SPARQL endpoints (e.g., https://www.w3.org/wiki/SparqlEndpoints)...

• Upload the dataset to a Linked Data catalog or repository;
  • The Linked Open Data Cloud http://lod-cloud.net/
  • Data Hub https://datahub.io/search
  • CKAN (open source) https://ckan.org/

• Validate datasets
  • Vapour, a Linked Data validator

Current Digital Repositories and Linked Data

- **CONTENTdm**
- Not inside the system, but work on exported data;

- **UIUC**. Challenges of Mapping Digital Collections Metadata to Schema.org (Lampron et al. 2016)


A UIUC Experiment: [http://imagesearch-test1.library.illinois.edu/cdm/ref/collection/motley-new/id/599](http://imagesearch-test1.library.illinois.edu/cdm/ref/collection/motley-new/id/599)

Project Site: [http://publish.illinois.edu/linkedspcollections/](http://publish.illinois.edu/linkedspcollections/)
Current Digital Repositories and Linked Data

- **Samvera** (previously “Hydra”)
  - LDP platform, publication of RDF data;
  - Extensions: GeoMash, ORCiD integration
  - Case: [Oregon Digital](https://oregondigital.org/)
    - Migrated from CONTENTdm to Hydra;
    - Created [opaquenamespace](https://schema.org/blanknode) (to accommodate terms without existing LOD links)

- **Islandora**
  - Effort in mapping MODS to RDF;
  - Fedora 4, LDP implementation; Islandora CLAW
  - Experiments and Cases:
    - [Islandora CLAW](https://www.ala.org/alsip/islandora-claw)
    - [TU Delft repository example: Colonial Architecture](https://delftdora.nl/) (Delftdora module, Fedora 3)
    - [A Lightweight Structured Data Implementation Using JSON-LD and Schema.org for Digital Repository](http://dx.doi.org/10.1109/TCC.2015.2410082) (Mak et al.)
Current Digital Repositories and Linked Data

• DSpace
  • DSpace Linked Data support spans all three Layers: “the storage layer with a triple store, the business logic with classes to convert stored contents into RDF, and the application layer with a module to publish RDF serializations.”
    - Donohue. Linked (Open) Data. (https://wiki.duraspace.org/display/DSDOC6x/Linked+%28Open%29+Data)

• Becker, Pascal-Nicolas. Illuminating DSpace's Linked Data Support
Current Digital Repositories and Linked Data

• Omeka
  - Open source web publishing platform. Supports Dublin Core. Can export METS, dc-rdf, JSON etc.
  - Ozmeka (plug-ins)/ Omeka linked data

• Item Relations (Plug-in)
  - It defines item-item relationship using several vocabularies (e.g., Dublin Core, FRBR, FOAF, and BIBO). For example,

**Title:** Linked data in the library

**Is Part Of:** CALA Southeast Chapter 2017 Spring Program "Sharing Your Professional Expertise and Your Library's Collections"

**Item Relations**

<table>
<thead>
<tr>
<th>This Item</th>
<th>Item: CALA Southeast Chapter 2017 Spring Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>frbr:partOf</td>
<td></td>
</tr>
<tr>
<td>dcterms:isPartOf</td>
<td></td>
</tr>
</tbody>
</table>
Current Digital Repositories and Linked Data

• Digital Commons
  • Allows adding extra fields for links (not RDF)

• Preparing for Linked Data
  • Include links (e.g., id.loc.gov, viaf.org) in added additional fields for author, advisor, department names etc.;

  • Data (in spreadsheet) cleanup and reconciliation using OpenRefine
Data Editing Tool: OpenRefine

- OpenRefine
  - Download and installation
    http://openrefine.org/download.html
  - Documentation
    https://github.com/OpenRefine/OpenRefine/wiki/Documentation-For-Users
- Features
  - Import
  - Facet
  - Edit cells: Clustering...
  - Edit columns: Add Column based on this column; Fetching URLs from Web Services (*Enrich existing data columns)...
  - Expressions
  - Export
  - History...

- Adding LOD Links to Fields in Digital Repositories
  - Authors/Creators
  - Contributors, e.g., thesis advisors
  - Corporate names, e.g., university, college and departmental names
  - Subjects
  - Person identifiers
Adding Linked Data Vocabularies to UCF Digital Repositories: Cases

• Cases
  • Adding LC and VIAF Links to Advisor Names
    • Faceting, clustering, LoC reconciliation, review; add column based on this column, VIAF reconciliation, review
  
  • Combining Multiple Links for Author Names
    • Expression: General Refine Expression Language (GREL)

  • Adding LC Subject Links to IR Records
    • Splitting, add column based on this column, expression

  • Adding FAST Headings by Fetching URLs
    • Add column by fetching URLs, expression, text transformation

• For step by step instructions, please refer to
  A Step Forward: Adding Linked Data Vocabularies to Digital Repositories
Adding LC and VIAF Links to Advisor Names

**Faceting, clustering;**
- Click “Reconcile—Start reconciliation”;
- Add and choose Reconciliation Service;
- Reconciliation & review;
- **Add/Display LoC URLs:**
  - Edit column—Add column based on this column
  - **New column name:** advisor_lc
  - **Expression:** cell.recon.match.id

---

**How to set up LoC Reconciliation Service**

Click “Add Standard Service”;

Enter the service’s URL:

https://lc-reconcile.herokuapp.com/
Adding LC and VIAF Links to Advisor Names

VIAF-LC Reconciliation Service
http://refine.codefork.com/reconcile/viaf/LC

Virtual International Authority File service
http://iphylo.org/~rpage/phyloinformatics/services/reconciliation_viaf.php
Adding LC and VIAF Links to Advisor Names

Add VIAF Links

New column name: advisor_viaf
Expression:
"https://viaf.org/viaf/" + cell.recon.match.id

VIAF links added
Adding LC and VIAF Links to Advisor Names

- Display in the IR

STARS Citation
Wanielista, Joseph R., "Removal of Color From Surface Water in Central Florida Retrospective Theses and Dissertations. 80.
http://stars.library.ucf.edu/rtd/80

Contributor (Linked data)
Yousef, Yousef A. [VIAF]
Yousef, Yousef A. [LC]
University of Central Florida, College of Engineering [VIAF]
University of Central Florida, College of Engineering [LC]
Collection (Linked data)
Retrospective Theses and Dissertations

*LOD links added for: Advisor name College name

*Adding LOD links for the advisor names is a work in progress.

STARS Citation
http://stars.library.ucf.edu/rtd/211

Contributor (Linked data)
Yousef, Yousef A.

University of Central Florida, College of Engineering
Collection (Linked data)
Retrospective Theses and Dissertations

http://stars.library.ucf.edu/rtd/211/

*Add both links or only one type of link for advisors (e.g., VIAF, or LC)
Combining Multiple Links for Author Names

• Using Expression to Format Multiple Links
  • Expression Language: General Refine Expression Language (GREL)
  • First, run reconciliations and get VIAF and LC links;
  • Then add columns, e.g.,
    • Edit column – Add column based on this column
    • New Column name  author1\_viafurl
    • Expression  "http://viaf.org/viaf/" + cell.recon.match.id
    • Edit column – Add column based on this column
    • New Column name  author1\_lcurl
    • Expression  cell.recon.match.id
Combining Multiple Links for Author Names

• Using Expression to Format Multiple Links
  • Add column based on this column
    • New column name: `author1_viafurl_lcurl`
    • Expression
      ```
      cells["author1_viafurl"].value + " " + cells["author1"].value + 
      "[VIAF] " + cells["author1_lcurl"].value + " " + 
      cells["author1_copy"].value + "[LC]"
      ```
Adding LC Subject Links to IR Records

Main Steps:
• Splitting keywords into columns;
• Edit column – Add column based on this column:
  New column name
  keyword1_url
  Expression
  cell.recon.match.id + " | " +
  cell.recon.match.name
  ...
• Put all subjects & links together in a column:
  New column name
  subjects_linked
  Expression
  if(isBlank(cells["keyword1_url"].value),
  " ", cells["keyword1_url"].value) + " "+
  ...
  if(isBlank(cells["keyword6_url"].value),
  " ", cells["keyword6_url"].value)....

<table>
<thead>
<tr>
<th>subjects_linked</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://id.loc.gov/authorities/childrensSubjects/sj96004925">http://id.loc.gov/authorities/childrensSubjects/sj96004925</a></td>
</tr>
<tr>
<td><a href="http://id.loc.gov/authorities/childrensSubjects/sj96006075">http://id.loc.gov/authorities/childrensSubjects/sj96006075</a></td>
</tr>
</tbody>
</table>
Adding LOD Links to IR Records

An Example in the IR
http://stars.library.ucf.edu/ucf-forum/186/
Adding FAST by Fetching URLs

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit column—add column by fetching URLs (under “Topical_Subject”):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New column name</td>
<td>fast_url</td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>&quot;<a href="http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22+value.replace(/(%5Cs)/,'%20')+%22&amp;httpAccept=application/rdf%2bxml%22">http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22+value.replace(/(\s)/,'%20')+%22&amp;httpAccept=application/rdf%2bxml&quot;</a></td>
<td></td>
</tr>
<tr>
<td>Text transformation (extract FAST ID):</td>
<td>Expression &quot;<a href="http://experimental.worldcat.org/fast/%22+value.parseHtml().select('dct">http://experimental.worldcat.org/fast/&quot;+value.parseHtml().select('dct</a></td>
<td>identifier')[0].htmlText()</td>
</tr>
</tbody>
</table>
Other Possibilities for Adding LOD Links to Digital Repositories and OpenRefine

• Augment Data with Reconciliation Services
  • Reconcile genre values to AAT terms
  • Reconcile places to GeoNames
  • Reconcile ORCID...

• Add Columns by Fetching URLs
  • Geocode your data using a service like Google map
  • Gathering IR seed data with OpenRefine and SHERPA/RoMEO

• Reconcile Against Local Vocabularies
  • reconciliation-csv
    • http://okfnlabs.org/reconcile-csv/
    • Possibilities: Reconcile various fields in OpenRefine against local controlled term lists, e.g., medals, ranks for Veterans Oral History

• Next Steps
  • Keep adding more LOD links to personal names, subjects and college/department names to digital collections;
  • Continue to explore OpenRefine and related services to enrich metadata and improve workflow;
  • Promote, explore and link authors/researchers to personal identifiers.
Reflections

• Possible Ways to Get Involved with Linked Data
  • Be aware of what’s going on, such as major LOD projects in the library community and the web;
  • Attend webinars and trainings to learn;
  • Be an advocate of open data;
  • Help faculty members and graduate students to understand linked data (such as linked data in general and in their disciplines), use open data and linked open data to support research;
  • Promote linked data and identity management (e.g., ORCID), make researchers aware of their presence in established linked data set (e.g., VIAF, LC LOD);
  • Collaborate and contribute to linked data projects when possible, e.g.,
    • Add LOD links to enrich record information and improve workflow using tools (e.g., OpenRefine);
    • Enrich project content by auto-annotating using tools (e.g., DBpediaSpotlight);
    • Support in applying and designing LOD thesauri and dataset;
    • Watch for LOD developments in state and national level initiatives;
    • Facilitate research in using linked open data...
Resources

- Becker, Pascal-Nicolas. **Illuminating DSpace's Linked Data Support**
- Berners-Lee, Tim. **Linked Data** [https://www.w3.org/DesignIssues/LinkedData.html](https://www.w3.org/DesignIssues/LinkedData.html)
- BIBFRAME [http://BIBFRAME.org/](http://BIBFRAME.org/)
- Creative Commons Rights Expression Language (CC REL) [https://wiki.creativecommons.org/wiki/CC_REL](https://wiki.creativecommons.org/wiki/CC_REL)
- Data Linkage Infrastructure - Census Bureau [https://www.census.gov/about/adrm/linkage/about.html](https://www.census.gov/about/adrm/linkage/about.html)
- DATA.GOV [https://www.data.gov/](https://www.data.gov/)
- DBpediaOntology Classes [http://mappings.DBpedia.org/server/ontology/classes/](http://mappings.DBpedia.org/server/ontology/classes/)
- DBpedia page: Jane Austen [http://DBpedia.org/page/Jane_Austen](http://DBpedia.org/page/Jane_Austen)
- Deng, Sai. **A Step Forward: Adding Linked Data Vocabularies to Digital Repositories.** [http://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1652&context=ucfscholar](http://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1652&context=ucfscholar)
- DPLA [http://dp.la](http://dp.la)
- Europeana [http://europeana.eu](http://europeana.eu)
- Facebook’s Open Graph [https://developers.facebook.com/docs/sharing/opengraph/](https://developers.facebook.com/docs/sharing/opengraph/)
- FAST: Faceted Application of Subject Terminology. [http://id.worldcat.org/fast](http://id.worldcat.org/fast)
- Google Knowledge Graph [https://en.wikipedia.org/wiki/Knowledge_Graph](https://en.wikipedia.org/wiki/Knowledge_Graph)
- International Standard Name Identifier (ISNI) [http://www.isni.org/](http://www.isni.org/)
- LC Linked Data Service: Authorities and Vocabularies [https://id.loc.gov/](https://id.loc.gov/)
- Library.Link Network [http://library.link/](http://library.link/)
- Mitchell, Erik T. **Building Blocks of Linked Open Data in Libraries. Library Technology Reports (vol. 49, no. 5).**
Resources

- NCSU Libraries: Organization Name Linked Data. http://lib.ncsu.edu/id/onlId
- OCLC Linked Data https://www.oclc.org/developer/develop/linked-data.en.html
- Open Library https://openlibrary.org/
- Open Researcher and Contributor ID (ORCID) https://orcid.org/
- Oregon Digital https://oregondigital.org/catalog/
- Ozmeka https://github.com/ozmeka/ozmeka
- RDA Registry http://www.rdaregistry.info/
- SHARE-VDE (Virtual Discovery Environment) http://www.share-vde.org/
- Simple Knowledge Organization Systems (SKOS) https://www.w3.org/2004/02/skos/
- SPARQL Explorer http://DBpedia.org/snorql/
- The Linked Open Data Cloud https://lod-cloud.net/
- University of North Texas Libraries. UNT Names. http://digital2.library.unt.edu/name
- Virtual International Authority File (VIAF) https://viaf.org/
- WorldCat Identities www.worldcat.org/identities
Acknowledgements

• Lucas Mak, Metadata and Catalog Librarian, Michigan State University Libraries

• Systems, University of Central Florida Libraries

• Image credits
  • Cover background: https://www.ebscohost.com/novelist-the-latest/blog-article/linked-data-what-is-it-and-why-should-you-care
  • Linked Data https://www.w3.org/DesignIssues/LinkedData.html
  • RDF Graph for RDA Class Hierarchies http://www.rdaregistry.info/rgAbout/rdaont/ontHierarchies.html
Thank you!

Contact
Sai Deng
Metadata Librarian & Associate Librarian
University of Central Florida Libraries
sai.deng@ucf.edu