Lower the Barrier & Be Empowered: Creating and Including Linked Data Vocabularies for Digital Collections

1-26-2019

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

Part of the Cataloging and Metadata Commons

Original Citation


This Conference Presentation is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
Lower the Barrier & Be Empowered: Creating and Including Linked Data Vocabularies for Digital Collections

Sai Deng, University of Central Florida Libraries
Digital Repositories and Linked Data

• **STARS** ([https://stars.library.ucf.edu](https://stars.library.ucf.edu))
  - UCF’s IR; Based on Digital Commons;
  - Allow to add extra fields for LOD links

• **Preparing for Linked Data**
  - Include links (e.g., viaf.org, id.loc.gov, Wikidata) in added additional fields for authors, advisors, college names, subjects etc.;

  • Data cleanup and reconciliation using OpenRefine
Digital Repositories and Linked Data

- **Islandora**
  - Effort in mapping MODS to RDF;
  - Fedora 4, LDP implementation; Islandora CLAW

- **Samvera** (previously “Hydra”)
  - LDP platform, publication of RDF data;
  - Case: Oregon Digital
    - Migrated from CONTENTdm to Hydra;
    - Created opaquenamespace

- **CONTENTdm**
  - Not inside the system, but work on exported data;
  - Challenges of Mapping Digital Collections Metadata to Schema.org: Working with CONTENTdm (Lampron et. al., 2016)
  - OCLC& CONTENTdm Metadata Refinery, transforms CONTENTdm metadata to Linked Data, 2016. Not yet scalable. (Washburn et al., 2016)
  - Linked Data Research - OCLC
Digital Repositories and Linked Data

• Omeka
  • Open source web publishing platform;
  • Supports DC, can export METS, dc-rdf, JSON etc.
  • Ozmeka (plug-ins)/ Omeka 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)
  • Illuminating DSpace's Linked Data Support (Becker, 2016)
Adding Linked Data Vocabularies to UCF Digital Repositories

• 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

• 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 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

Open OpenRefine — Choose Files — Select rtd.xls — Click “Next”
Adding LC and VIAF Links to Advisor Names

Click “Create Project”
Adding LC and VIAF Links to Advisor Names

First click “Facet – Text facet” (under advisor1)
Adding LC and VIAF Links to Advisor Names

Clustering

Click “Cluster” on the left, or click “Edit Cells – Cluster and edit” (under “advisor1” column)
Adding LC and VIAF Links to Advisor Names

Cluster & Edit column “advisor1”
(Method: key collision, Keying Function: fingerprint)

- Hynes, Michael C. (45 rows)
- Hynes, Michael C.
- Hynes, Michael C. (3 rows)
  ➔ Hynes, Michael C.

- Chow, Lee (7 rows)
  ➔ Chow, Lee
- Lee, Chow (1 rows)
  ➔ Chow, Lee
Adding LC and VIAF Links to Advisor Names

Cluster & Edit column “advisor”:
Try other algorithms; Merge Selected & Close

Might be hard to determine at early stage:
Butler, Jeff (8 rows)
Butler, Jeff F. (3 rows)

Gay, David A. (3 rows)
Gay, David (1 rows)

(*Appearance time? Middle name? What’s established?)
Adding LC and VIAF Links to Advisor Names

Reconciliation

Sort by advisor;

Click “Reconcile—Start reconciliation”
Adding LC and VIAF Links to Advisor Names

How to set up LoC Reconciliation Service

Click “Add Standard Service”;

Enter the service's URL:

https://lc-reconcile.herokuapp.com/

Add Service
Adding LC and VIAF Links to Advisor Names

Reconciliation

Leave “Names” as default;
Click “Start Reconciling”

Matched 1549 (about 42.44%).
For names linked, click the name to check;
If it’s not the right person, click “Choose new match” for other options.

Adams, Sean P.
URI(s) [http://id.loc.gov/authorities/names/n92070159](http://id.loc.gov/authorities/names/n92070159)

Sources
found: Letters from the front, 1898-1945, 1992:CIP t.p. (Sean P. Adams)
found: Old Dominion, industrial commonwealth coal, politics, and economy in antebellum America, 2004:ECIP t.p. (Sean Patrick Adams) dataview (b. 1969; Department of History, Univ. of Central Fla.)
Adding LC and VIAF Links to Advisor Names

Review

• For names unlinked, click the options below to check;

• If it’s the right person, click ✓ (apply to this cell only) or ✓✓ (apply to all identical cells) to apply.

• If no right name found, click “Create new topic”;

• Input the right name and apply.

Click the link under “Chase, Diane Z.”, it opens: http://id.loc.gov/authorities/names/n89612783

Sources
found: nuc89-21890: Her Offerings to the gods, 1986(hdg. on DDO rept.: Chase, Diane Z.; usage: Diane Z. Chase)
found: Her A postclassic perspective, 1988:t.p. (Diane Zaino Chase) p. 1 (Univ. of Central Florida)
Adding LC and VIAF Links to Advisor Names

Review

- Can be time-consuming;
- Determining factors: name form, institution, field, book/work info, birth place, year...
- When LCNAF doesn’t provide enough information, further check the web;
- When confirmed that it’s the right person, click ✓ or ✓✓ to apply.

Hubbard, Susan
URI(s) http://id.loc.gov/authorities/names/n88050761

Sources
found: Her Is repeating a grade a sign of failure? 1988:p. 6...
found: Phone call to S. Hubbard 5-14-90 (Susan Hubbard, author of Is repeating a grade ... is the same person; b. 9-6-51)

Hubbard, Susan 1951- (Susan Mary Hubbard) (Encyclopedia)
Born September 6, 1951, in Syracuse, NY... Home—Winter Park, FL. Office—Department of English, University of Central Florida
Adding LC and VIAF Links to Advisor Names

Lillios, Anna, 1948-
URI(s) http://id.loc.gov/authorities/names/n2003044537
Sources
found: Lawrence Durrell and the Greek world, c2004:EDCIP t.p. (Anna Lillios) data view (b. 3/13/48)
Google books: Lawrence Durrell and the Greek World
About the author (2004)
Anna Lillios, associate professor of English at the University of Central Florida, is the editor of the "Marjorie Kinnan Rawlings Journal of Florida Literature“...

Lue, Martha Scott
URI(s) http://id.loc.gov/authorities/names/n00091486
Sources
found: Lue, Martha Scott. A survey of communication disorders for the classroom teacher, 2000:CIP t.p. (Martha Scott Lue)
CV in UCF Website: Martha S. Lue Stewart

Review

- What can be checked:
  - Google
  - Wikipedia
  - Amazon books
  - Publisher site
  - Obituaries
  - University website
  - Publications
  - CVs
  - Encyclopedia
  - Newsletter
  - Archives
  - ...
Adding LC and VIAF Links to Advisor Names

Add Column for URLs

- Edit column—Add column based on this column (under advisor1)

- In the opened new window, fill in:
  New column name: advisor1_lc
  Expression: cell.recon.match.id

LC links added to the data sheet
Adding LC and VIAF Links to Advisor Names

Add VIAF Links

Add a duplicate column for advisor1: advisor1_c.

Add a column based on advisor1_c:
New column name: advisor1_viaf
Expression: 
"https://viaf.org/viaf/" + cell.recon.match.id

[advisor1_viaf] Reconciliation
Select VIAF-LC Reconciliation Service (Check LC names only)
(*or Virtual International Authority File service)

How to set up VIAF-LC Reconciliation Service
Click “Add Standard Service”;
Enter the service’s URL: http://refine.codefork.com/reconcile/viaf/LC

How to set up Virtual International Authority File service
Enter the service’s URL: http://iphylo.org/~rpage/phyloinformatics/services/reconciliation_viaf.php
Adding LC and VIAF Links to Advisor Names

Combine Values

Using Expression to Format Multiple Links

- Add column based on this column
  - New column name: advisor1_viaf_lc
  - Expression

If there are no blank cells:

```grel
if(isBlank(cells["advisor1_viaf"].value), " ",
cells["advisor1_viaf"].value + " " +
cells["advisor1"].value + "[VIAF] " +
cells["advisor1_lc"].value + " " +
cells["advisor1_c"].value + "[LC]"
```

If there are blank cells:

```grel
if(isBlank(cells["advisor1_viaf"].value), " ",
cells["advisor1_viaf"].value + " " +
if(isBlank(cells["advisor1_c"].value), " ",
cells["advisor1_c"].value) + "[VIAF] " +
if(isBlank(cells["advisor1_lc"].value), " ",
cells["advisor1_lc"].value) + " " +
if(isBlank(cells["advisor1"].value), " ",
cells["advisor1"].value) + "[LC]"
```
An Example in the IR (with VIAF/LC links)

STARS Citation
https://stars.library.ucf.edu/rtd/266

Contributor (Linked data)
McLellon, Waldron M. [VIAF]

McLellon, Waldron M. [LC]

University of Central Florida, College of Engineering [VIAF]

University of Central Florida, College of Engineering [LC]

Collection (Linked data)
Retrospective Theses and Dissertations

https://stars.library.ucf.edu/rtd/266/
Adding Locally Created Wikidata for RTDs

- Wikidata is created for local terms.

**STARS Citation**
https://stars.library.ucf.edu/rtd/3669

**Contributor (Linked data)**
Richard R. Adicks (Q57704630)

**STARS Citation**
https://stars.library.ucf.edu/rtd/251

**Contributor (Linked data)**
Christian S. Bauer, Jr. (Q57708477)

**Collection (Linked data)**
*Retrospective Theses and Dissertations*

LOD links added for Advisor name, e.g., Richard R. Adicks (Q57704630) https://www.wikidata.org/wiki/Q57704630
https://stars.library.ucf.edu/rtd/3669/
https://stars.library.ucf.edu/rtd/251/
Adding Locally Created Wikidata for RTDs

Richard R. Adicks (Q57704630)
https://www.wikidata.org/wiki/Q57704630

Statements
- instance of
- sex or gender
- given name

Properties
- instance of
- sex or gender
- given name

Create a new Item
Adding Locally Created Wikidata for RTDs

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
<th>References</th>
<th>Edit Link</th>
<th>Add Reference</th>
<th>Add Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>family name</td>
<td>Adicks</td>
<td>0</td>
<td>[edit]</td>
<td>[add reference]</td>
<td>[add value]</td>
</tr>
<tr>
<td>place of birth</td>
<td>Lake City</td>
<td>1</td>
<td>[edit]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date of death</td>
<td>15 August 2011</td>
<td>1</td>
<td>[edit]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>place of death</td>
<td>Lakeland</td>
<td>1</td>
<td>[edit]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>occupation</td>
<td>university teacher</td>
<td>0</td>
<td>[edit]</td>
<td>[add reference]</td>
<td>[add value]</td>
</tr>
<tr>
<td>field of work</td>
<td>English</td>
<td>0</td>
<td>[edit]</td>
<td>[add reference]</td>
<td>[add value]</td>
</tr>
</tbody>
</table>

Richard R. Adicks (Q57704630)

Statements

... family name place of birth date of death place of death occupation field of work ...

Property Reference

Reference URL: https://www.legacy.com/obituaries/orlandosentinel/obituary.aspx?n=richard-adicks&pid=175526324
Adding Locally Created Wikidata for RTDs

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>employer</td>
<td>University of Central Florida</td>
</tr>
<tr>
<td></td>
<td>Rollins College</td>
</tr>
<tr>
<td></td>
<td>Georgia Institute of Technology</td>
</tr>
<tr>
<td>award received</td>
<td>UCF Faculty Excellence in Teaching Award</td>
</tr>
<tr>
<td>educated at</td>
<td>Tulane University</td>
</tr>
<tr>
<td>academic degree</td>
<td>Doctor of Philosophy</td>
</tr>
<tr>
<td>notable work</td>
<td>A Court for Owls</td>
</tr>
<tr>
<td></td>
<td>Oviedo: Biography of a Town</td>
</tr>
</tbody>
</table>

Richard R. Adicks (Q57704630)

Statements

Properties

- employer
- award received
- educated at
- academic degree
- notable work

Book Name
- instance of written work

written work (Q47461344)
- subclass of work
STARS Citation
https://stars.library.ucf.edu/rtd/499

Contributor (Linked data)
Hosni, Djehane, A. [VIAF]
Hosni, Djehane, A. [LC]

University of Central Florida, College of Business Administration (Q7895232)

Collection (Linked data)
Retrospective Theses and Dissertations

https://stars.library.ucf.edu/rtd/499/

University of Central Florida. College of Business Administration (Q7895232)
https://www.wikidata.org/wiki/Q7895232
A Special Example with Multiple LOD Links: Wikidata, VIAF, LC

STARS Citation
https://stars.library.ucf.edu/rtd/4818

Contributor (Linked data)
Faissal A. Mosleh (Q59495035)
Mosleh, Faissal Abdel Haleem [VIAF]
Mosleh, Faissal Abdel Haleem [LC]
University of Central Florida. College of Engineering [VIAF]
University of Central Florida. College of Engineering [LC]

*Normally we only create Wikidata entries for names that are not in the VIAF or LC database.
Faissal A. Moslehy (Q59495035)

Statements
- instance of
- sex or gender
- given name
- family name
- occupation
- field of work
- employer
- member of
- award received
- educated at
- academic degree
- official website
<table>
<thead>
<tr>
<th>Identifiers</th>
<th>VIAF ID</th>
<th>Library of Congress authority ID</th>
<th>ISNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIAF ID</td>
<td>96900524</td>
<td>n82027354</td>
<td>0000 0000 6895 0732</td>
</tr>
<tr>
<td>Library of Congress authority ID</td>
<td>0 references</td>
<td>1 reference</td>
<td></td>
</tr>
<tr>
<td>ISNI</td>
<td>0000 0000 6895 0732</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISNI</td>
<td>0000 0000 6895 0732</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reference stated in VIAF ID reference URL

Virtual International Authority File
Querying Wikidata
Querying Wikidata: Graphical Presentations

Graph: University Teachers, Gender and Fields

*The data pool is very small. Not representative as of now.*
Querying Wikidata: Graphical Presentations

Scatter Chart: University Teachers, Gender and Fields

*The data pool is very small. Not representative as of now.*
Discussion

• Possible Ways to Get Involved with Linked Data
  • Be aware of what’s going on, attend webinars/trainings;
  • Be an advocate of open data;
  • Help researchers to understand linked data, use open data and linked data to support research;
  • Promote linked data and identity management (e.g., ORCID);
  • Collaborate and contribute to linked data projects when possible, e.g., add LOD links to enrich records information...

• Recommended Wikidata Practices for Libraries to be Determined
  • Creating local authority records using Wikidata platform is a low barrier linked data practice;
  • For terms already in the national or international authority databases, is it better for local libraries not to add Wikidata entries even they can be linked to VIAF/LC and additional information is available? (What about OCLC’s Wikibase experiment?)
  • If Wikidata entries are added by libraries for terms already in the VIAF/LC authority database, should we wait for the Wikidata’s identifiers/external links (e.g., VIAF/LC/ISNI etc.) to be imported by national efforts?
  • What’s the current status in authority control with various methods by different entities?

• For us, we will continue to explore OpenRefine, Wikidata and related services to enrich metadata and improve workflow for our digital collections.
Resources

- Berners-Lee, Tim. Linked data. [https://www.w3.org/DesignIssues/LinkedData.html](https://www.w3.org/DesignIssues/LinkedData.html)
- Becker (2016). Illuminating DSpace’s linked data support. [https://www.slideshare.net/pascalbecker/illuminating-dspaces-linked-data-support](https://www.slideshare.net/pascalbecker/illuminating-dspaces-linked-data-support)
- Deng (2018). A step forward: Adding linked data vocabularies to digital repositories. [https://stars.library.ucf.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1652&context=ucfscholar](https://stars.library.ucf.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1652&context=ucfscholar)
- DSpace: Linked (open) data. [https://wiki.duraspace.org/display/DSDOC6x/Linked+%28Open%29+Data](https://wiki.duraspace.org/display/DSDOC6x/Linked+%28Open%29+Data)
- FAST: Faceted Application of Subject Terminology. [http://id.worldcat.org/fast](http://id.worldcat.org/fast)
- Islandora CLAW. [https://islandora.ca/CLAW](https://islandora.ca/CLAW)
- LC Linked Data Service: Authorities and vocabularies. [https://id.loc.gov/](https://id.loc.gov/)
- Open Researcher and Contributor ID (ORCID). [https://orcid.org/](https://orcid.org/)
Resources

- Ozmeka. https://github.com/ozmeka/ozmeka
- Reconcile-csv: http://okfnlabs.org/reconcile-csv/
  Reconcile-csv in GitHub: https://github.com/okfn/reconcile-csv
- The Linked Open Data Cloud. https://lod-cloud.net/
- Virtual International Authority File (VIAF). https://viaf.org/
- Wikidata. https://www.wikidata.org
- Worldcat Identities. www.worldcat.org/identities
Acknowledgement
Information Technology & Digital Initiatives
University of Central Florida Libraries
Thank you!

Contact

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