Making new discoveries from old data: Utilizing digital scholarship to foster new research in Special Collections

Carruthers, Matthew

http://hdl.handle.net/2027.42/138108
Making new discoveries from old data

Utilizing digital scholarship to foster new research in special collections

Slides: goo.gl/smvkvf
Goals

1. Explore using digital scholarship methods and tools to enhance user experience in special collections and potentially provide new services.

2. Develop a pilot project as a proof of concept for a new service.
   - Utilize existing open source technologies.
   - Little to no need for dedicated IT support.
Result

Provide visualization on demand, customized to a patron’s individual research question.

- Use data from finding aids and publicly accessible databases along with open-source tools to create interactive visualizations of social networks of the creators of our archival collections.
- Make connections between entities more apparent and provide a discovery tool to spur further inquiry and research.
Data Processing

Encoded Archival Description (EAD)
- Metadata standard for the encoding of finding aids for use in a networked (online) environment.

Encoded Archival Context – Corporate Bodies, Persons, and Families (EAC-CPF)
- Companion standard to Encoded Archival Description (EAD).
- Describes the entity responsible for creating (or who is represented in) one or more archival collections.
Data Processing: Tools

EAC-CPF record creation:

Social Networks and Archival Context
http://socialarchive.iath.virginia.edu/snac/search

Remixing Archival Metadata Project
http://rampeditor.info

xEAC
https://github.com/ewg118/xEAC
More Data Processing

Use XSLT to extract information about connections and format output files as tab-delimited text.
More Data Processing

Information extracted:
• Name of connection
• Type of entity
• Type of connection
• Link to more information (when available)

```xml
<cpfRelation
 xlink:arcrole="http://socialarchive.iath.virginia.edu/control/term#associatedWith"
xlink:href="http://n2t.net/ark:/99166/w6f48mrk"
xlink:role="http://socialarchive.iath.virginia.edu/control/term#Person"
xlink:type="simple" xmlns:xlink="http://www.w3.org/1999/xlink">
<relationEntry>Zirlin, Larry.</relationEntry>
</cpfRelation>
```
More Data Processing

Information extracted:

- Name of connection
- Type of entity
- Type of connection
- Link to more information (when available)
More Data Processing

Information extracted:
- Name of connection
- Type of entity
- Type of connection
- Link to more information (when available)

Entity types:
- Person
- Corporate Body
- Family
- Archival Resource
- Bibliographic Resource

```xml
<cpfRelation
  xlink:arcrole="http://socialarchive.iath.virginia.edu/control/term#associatedWith"
  xlink:href="http://n2t.net/ark:/99166/w6f48mrk"
  xlink:role="http://socialarchive.iath.virginia.edu/control/term#Person"
  xlink:type="simple" xmlns:xlink="http://www.w3.org/1999/xlink">
  <relationEntry>Zirlin, Larry.</relationEntry>
</cpfRelation>
```
The Process

Information extracted:
- Name of connection
- Type of entity
- Type of connection
- Link to more information (when available)

Connection types:
- Associated With
- Corresponded With
- Creator Of
- Referenced In

```xml
<cpfRelation>
  xlink:arcrole="http://socialarchive.iath.virginia.edu/control/term#associatedWith"
  xlink:href="http://n2t.net/ark:/99166/w6f48mrk"
  xlink:role="http://socialarchive.iath.virginia.edu/control/term#Person"
  xlink:type="simple" xmlns:xlink="http://www.w3.org/1999/xlink">
    <relationEntry>Zirlin, Larry.</relationEntry>
  </cpfRelation>
```
More Data Processing

Information extracted:

- Name of connection
- Type of entity
- Type of connection
- Link to more information (when available)

```xml
<cpfRelation>
    xlink:arcrole="http://socialarchive.lauch.virginia.edu/control/term#associatedWith"
    xlink:href="http://n2t.net/ark:/99166/w6f48mrk"
    xlink:role="http://socialarchive.lauch.virginia.edu/control/term#Person"
    xlink:type="simple" xmlns:xlink="http://www.w3.org/1999/xlink">
        <relationEntry>Zirlin, Larry.</relationEntry>
    </cpfRelation>
```
Visualizing the Data

Cytoscape – http://cytoscape.org/

- Open source desktop application for visualizing complex networks.
### Edge Color (Unselected) Mapping

<table>
<thead>
<tr>
<th>Edge Color (Unselected)</th>
<th>interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Green" /></td>
<td>associatedWith</td>
</tr>
<tr>
<td><img src="image" alt="Yellow" /></td>
<td>correspondedWith</td>
</tr>
<tr>
<td><img src="image" alt="Blue" /></td>
<td>creatorOf</td>
</tr>
<tr>
<td><img src="image" alt="Pink" /></td>
<td>referencedIn</td>
</tr>
</tbody>
</table>
Sharing the Visualizations

Share visualizations via web-publishing companion platform to Cytoscape:

CyNetShare

http://cynetshare.ucsd.edu/#/

Working example:

http://goo.gl/eHYLUi
Sharing the Visualizations

CyNetShare

An easy way to share your network visualizations

- Project Web Site
- Documentation

- Public Files
- Gists

- Network Data URL
- Enter Network File URL

- Visual Style URL
- Enter Style URL

- Visualize!
Sharing the Visualizations

goo.gl/eHYLUi
Next Steps

Continue to streamline the workflow
Next Steps

Perform user testing
(Incorporate feedback into service model)
Next Steps

Explore more options for using and manipulating the data

(Customize and host local CyNetShare instance)

http://www-personal.umich.edu/~mcarruth/HanumanBooksVis/web_session/index.html#/

Code4Lib 2016
March 9, 2016
Thanks!

Matt Carruthers
mcarruth@umich.edu
@mattadata2

Slides: goo.gl/smvkvf