Integrating LOD into Library’s Digitized Special Collections

Myung-Ja K. Han (mhan3@Illinois.edu)
Deren Kudeki (dkudeki@illinois.edu)
Timothy W. Cole (t-cole3@Illinois.edu)
Jacob Jett (jjett2@Illinois.edu)
Introduction
Project Context

• Exploring the Benefits for Users of LOD for Digitized Special Collections
  - 18-month exploratory study
  - Funded by the Andrew W. Mellon Foundation

• Digitized Library Special Collections
  - Many relegated to information silos largely disconnected from the broader Web
  - How can we better connect these special resources to the Web?
    - Can we use Linked Open Data to help us? If so, how hard is it to do?

• Objectives
  - Map legacy metadata schemas to LOD-compliant schemas
  - Actively link to and from DBpedia, VIAF, wikidata, and related Web resources
Collections Tested

• The Motley Collection of Theatre & Costume Design
  - About 5,000 images of costume and set designs, sketches, production notes, and similar objects
  - Represents a variety of objects from the Motely Group’s career (1932-1976)

• Portraits of Actors, 1720-1920
  - Nearly 3,500 pictures of actors, including Sarah Siddons, Edmund Kean, and others

• Kolb-Proust Archive for Research
  - About 8,700 of Professor Philip Kolb’s research notecards on Marcel Proust
    - A chronology of events concerning Proust’s life
    - A bibliography of works mentioned in Proust’s correspondences
Schema.org as a Vehicle for Discovery

• Industry-wide use by Web search engines
• Some promising schema’s (e.g., Bibframe 2.0, etc.) were still under development at the time of the project’s beginning
• Some existing schema’s were considered to “heavy-weight” for the project’s data needs and goals (e.g., FRBR_{oo}, CIDOC-CRM, etc.)
• Some existing schema’s did not have wide-spread adoption (e.g., the SPAR family of ontologies)
• Were able to reuse previous library-oriented work (at UIUC and OCLC) with Schema.org
Collections -
1. Motley Collection of Theatre and Costume Design (Portraits of Actors, 1720-1920)
2. Kolb-Proust Archive Collection
Motley Collection of Theatre and Costume Design (New)

About this collection

The Motley Collection of Theatre and Costumes is housed at the Rare Book and Manuscript Library. It is a rich collection of materials related to theatre and costume design, including rare books, manuscripts, and original designs.

The collection includes over 500 items, ranging from early 19th-century designs to contemporary works. The collection is constantly being added to with new acquisitions and donations.

About this collection

Portraits of Actors, 1720-1920, includes almost 3500 pictures of actors' studio portraits and actors posing in costume for a particular role or performing a scene from a play. Dramatists, theatrical managers, singers and musicians are also included, but the majority are British and American actors who worked between about 1770 and 1893. Among the hundreds of actors included are: Sarah Siddons, Edmund Kean, John Philip Kemble, Edwin Booth, Edwin Forrest, William Henry West Betty, Charles Mathews, Dorothy Jordan, Frances Abington, and Ada Keane. The images were digitized from engravings, lithographs, mezzotints, aquatints, wood engravings, photographs, and photomechanically-reproduced prints, all from the University of Illinois Theatre Print Collection.
### Description

**1914: Sergeant and Grocer**

**View Description**

<table>
<thead>
<tr>
<th>Rating</th>
<th>★★★★★ Based on 0 rating(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Title</td>
<td>1914: Sergeant and Grocer</td>
</tr>
<tr>
<td>Performance Title</td>
<td>Unknown Soldier and His Wife</td>
</tr>
<tr>
<td>Author / Composer</td>
<td>Listov, Peter</td>
</tr>
<tr>
<td>Theater</td>
<td>Vivian Beaumont Theater (New York, N.Y.)</td>
</tr>
<tr>
<td>Opening Performance Date</td>
<td>1967</td>
</tr>
<tr>
<td>Object</td>
<td>Costume rendering</td>
</tr>
<tr>
<td>Type</td>
<td>Image</td>
</tr>
<tr>
<td>Material/Techniques</td>
<td>Felt tip pen</td>
</tr>
<tr>
<td>Dimensions</td>
<td>11 3/4 x 8 1/4</td>
</tr>
<tr>
<td>Notes</td>
<td>This production also played in Lyons, London, and Stratford (Connecticut) - the same designs were used for later productions. The actors' names on the designs do not match the cast list of the original production because they were inserted on the designs later, in pencil, and indicate a later cast. Notes on some of the designs are in French, perhaps added during the Lyons production, &quot;Solde Inconnu&quot;. Some costumes are dated &quot;Motley 72&quot;, and may have been new renderings prepared specifically for a later production, as some of the facts under 73011j are the same title. These three designs (Wife, Arab and Barbarian) belong to this batch, two of them being dated &quot;Motley 68&quot;</td>
</tr>
<tr>
<td>Subject I (AAT)</td>
<td>costume design</td>
</tr>
<tr>
<td>Subject II (TGM)</td>
<td>costume design (character dress)</td>
</tr>
<tr>
<td>Subject I (LCSH)</td>
<td>Theatrical productions</td>
</tr>
<tr>
<td>Subject II (LCSH)</td>
<td>Costume design drawings</td>
</tr>
<tr>
<td>Rights</td>
<td>Rare Book and Manuscript Library, University of Illinois at Urbana-Champaign. For questions regarding these materials, please contact the Rare Book and Manuscript Library: <a href="https://illinois.edu/irr/soe/7016277">https://illinois.edu/irr/soe/7016277</a>. For more information on the Motley Collection, please contact the Rare Book and Manuscript Library.</td>
</tr>
<tr>
<td>Physical Location</td>
<td>University of Illinois at Urbana-Champaign. Rare Book and Manuscript Library</td>
</tr>
<tr>
<td>Inventory Number</td>
<td>67076-057</td>
</tr>
<tr>
<td>JPEG2000 URL</td>
<td><a href="http://verona.grainer.illinois.edu/motley/jp2/670766_057jp2/">http://verona.grainer.illinois.edu/motley/jp2/670766_057jp2/</a></td>
</tr>
<tr>
<td>Collection Title</td>
<td>Motley Collection of Theatre and Costume Design (University of Illinois at Urbana-Champaign Library)</td>
</tr>
</tbody>
</table>
Performance Title: Unknown Soldier and His Wife
Author / Composer: Ustinov, Peter
Theater: Vivian Beaumont Theater (New York, N.Y.)
Opening Performance Date: 1967

Notes:
This production also played in Lyons, London, and Stratford (Connecticut) - the names on the designs do not match the cast list of the original production because they indicate a later cast. Notes on some of the designs are in French, perhaps added due to costumes are dated "Motley 72", and may have been new renderings prepared for 730111 and the same title. These three designs (Wife, Arab and Barbarian) by...
<schema:VisualArtWork>
  <schema:name>1914: Sergeant and Grocer</schema:name>
  <schema:genre>Costume rendering</schema:genre>
  <schema:isPartOf>
    <schema:Book>
      <schema:name>Unknown Soldier and …</schema:name>
      <schema:author>http://viaf.org/viaf/98273667</schema:author>
      <schema:sameAs>http://theatricalia.com/person/r85/peter-ustinov</schema:sameAs>
    </schema:Book>
    <schema:CreativeWork>(StageWork)</schema:CreativeWork>
  </schema:isPartOf>
</schema:VisualArtWork>
@type: "CreativeWork",
aditionalType: "scp:StageWork",
exampleOfWork: "Unknown Soldier and His Wife",
@type: "Book",
sameAs: [
],
@id: "https://en.wikipedia.org/wiki/The_Unknown_Soldier_and_His_Wife",
@type: "Person",
dateCreated: "1967",
locationCreated: "http://viaf.org/viaf/140952057",
sameAs: [
"http://theatricalia.com/person/r85/peter-ustinov",
"https://en.wikipedia.org/wiki/Peter_Ustinov"
]

@type: "VisualArtwork",
name: "1914: Sergeant and Grocer",
genre: "Costume rendering",
artform: "Image",

@type: "CreativeWork",
additionalType: "scp:StageWork",
exampleOfWork: "Unknown Soldier and His Wife",
@type: "Book",
sameAs: [
],
@id: "https://en.wikipedia.org/wiki/The_Unknown_Soldier_and_His_Wife",
@type: "Person",
dateCreated: "1967",
locationCreated: "http://viaf.org/viaf/140952057",
sameAs: [
"http://theatricalia.com/person/r85/peter-ustinov",
"https://en.wikipedia.org/wiki/Peter_Ustinov"
]
Metadata for Motley Collection

• Metadata structure is flat
• Metadata describes more than one ‘object’
• Element name includes contextual information
• Multiple values can be included in a single element
• Use a specialized/local controlled vocabulary
Collections –

1. Motley Collection of Theatre and Costume Design (Portraits of Actors, 1720-1920)

2. Kolb-Proust Archive Collection
1913 14° novembre

Bibesco (prince Antoine) (1878-1951)

1913 12 novembre Mercredi soir

Bichet à Cailloux pour faire annoncer Semma, le plaignant: "...mes derniers articles avaient pour vous avoir été à charge, et bien que j'ai renoncé à finir la série [en 1913]. D'autre part, H. Chevassu n'a pas pris ceux que j'avais envoyés cette année (ils étaient trop longs) pour le Supplement. C'est pourquoi je me suis contenté d'envoyer aux Annales et au Tasse des extraits qui y paraîtront."

Hahn (Reynaldo) (1874-1947)
1913 mercredi soir 12 novembre

Proust écrit à Calmette pour faire paraître, éplorant, qu’ils étaient trop longs pour le Supplément et au Temps des extraits qui y paraîtront.

Proust, Marcel; pseud. Horatio
Calmette, Gaston

1913 Cor XII, p. 308, n. 142 [Le mercredi soir 12 novembre 1913]
Proust écrit à Calmette pour faire annoncer Swann, se plaignant: "...mes derniers articles avaient paru vous avoir été à charge, si bien que j'ai renoncé à finir la série [en 1912]. D'autre part, M. Chevassu n'a pas pris ceux que j'avais envoyés cette année (ils étaient trop longs) pour le Supplément. C'est pourquoi je me suis contenté d'envoyer aux Annales et au Temps des extraits qui y paraitront."

à Gaston Calmette, Cor XII, p. 308, n. 142 [Le mercredi soir 12 novembre 1913]

Record:069410
## Partial Mapping of TEI Document Elements

<table>
<thead>
<tr>
<th>TEI</th>
<th>Schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>div1 @id</td>
<td>schema:Dataset</td>
</tr>
<tr>
<td></td>
<td>schema:author <a href="http://viaf.org/44300868">http://viaf.org/44300868</a></td>
</tr>
<tr>
<td>-&gt;head-&gt;date @value</td>
<td>schema:temporalCoverage [schema:DateTime]</td>
</tr>
<tr>
<td>-&gt;div2-&gt;p-&gt;name</td>
<td>schema:mentions [schema:Person]</td>
</tr>
<tr>
<td>-&gt;div2-&gt;note-&gt;name</td>
<td></td>
</tr>
<tr>
<td>-&gt;div2-&gt;p-&gt;title</td>
<td>schema:mentions [schema:CreativeWork]</td>
</tr>
<tr>
<td>-&gt;div2-&gt;note-&gt;title</td>
<td></td>
</tr>
<tr>
<td>-&gt;div2-&gt;(listBibl)-&gt;bibl</td>
<td>schema:citation [schema:CreativeWork]</td>
</tr>
</tbody>
</table>
Encoding Name Database

<table>
<thead>
<tr>
<th>Full Name</th>
<th>KeyCode</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daudet, Léon</td>
<td>daudet1</td>
<td>1868-1942, fils aîné d'Alphonse Daudet</td>
</tr>
<tr>
<td>Daudet, Marthe Allard, Mme Léon;</td>
<td>daudet6</td>
<td>1878-1960, cousine et 2ème femme de Léon Daudet, mariée en 1903</td>
</tr>
<tr>
<td>pseud. Pampille</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daudet, Philippe</td>
<td>daudet10</td>
<td>?-1923, fils de Léon Daudet</td>
</tr>
<tr>
<td>Daudet, Claire-Antoinette</td>
<td>daudet11</td>
<td>1918- ; fille de Marthe (née Allard) et Léon Daudet (LJP)</td>
</tr>
</tbody>
</table>

- schema:familyName
- schema:givenName
- schema:birthDate
- schema:deathDate
- schema:gender
- schema:nationality
- schema:knows
- schema:spouse
- schema:children
- schema:parent
- schema:sibling
- schema:relatedTo
- schema:jobTitle
Daudet, Marthe Allard (daudet6) -- 1878-1960, cousine et 2ème femme de Léon

<table>
<thead>
<tr>
<th>Full Name</th>
<th>KeyCode</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daudet, Léon</td>
<td>daudet1</td>
<td>1868-1942, fils aîné d'Alphonse Daudet</td>
</tr>
<tr>
<td>Daudet, Marthe Allard, Mme Léon;</td>
<td>daudet6</td>
<td>1878-1960, cousine et 2ème femme de Léon Daudet, mariée en 1903</td>
</tr>
<tr>
<td>pseud. Pampille</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daudet, Philippe</td>
<td>daudet10</td>
<td>1878-1923, fils de Léon Daudet</td>
</tr>
<tr>
<td>Daudet, Claire-Antoinette</td>
<td>daudet11</td>
<td>1918- ; fille de Marthe (née Allard) et Léon Daudet (LJP)</td>
</tr>
</tbody>
</table>
Mapping Challenges for Special Collections

• Target vocabulary (Schema) still missing some key entities
  • Specifically no way to differentiate the production of a play from the individual performances
  • Solved by locally extending Schema

• Many entities are not currently listed in linked data sources
  • For Kolb-Proust we assigned URIs to every name and then linked the ones listed in authority control databases to those databases
  • Could do this for other collections
Metadata Enrichment and Reconciliation Work
Export metadata

Review **element names** and **values**

Identify and perform metadata enhancement/reconciliation work with linked data sources* and authority data

Mapped local elements to Schema.org and ingested into the system

*Sources used for the process include Library of Congress (LC) Name Authority Files, Virtual International Authority Files (VIAF), Internet Movie Database (IMDb), Internet Broadway Database (IBDb), Wikipedia, Worldcat Identities, Theatricalia, and many more.
Sources Consulted in Manual Process

Sources Supporting Linked Data

- Library of Congress (LC) Name Authority Files
- Virtual International Authority File (VIAF)
- Internet Movie Database (IMDb)
- Internet Broadway Database (IBDb)
- Wikipedia
- Worldcat Identities

Other Web Resources

- Canadian Theatre Encyclopedia
- Encyclopedia Britannica
- Turner Classic Movies
- Goodreads
- Obituaries in various digital newspapers
- Australian Dictionary of Biography
- doollee.com
- Opera Scotland
- Copies of text on Amazon Books
- Theatricalia
Metadata Enrichment

• Providing Linking or Canonical URIs for
  ❑ Persons
    o E.g., Peter Ustinov, Marcel Proust, etc.
  ❑ Venues
    o E.g., the Old Victoria Theatre, Alexandra Theatre, etc.
  ❑ Plays/Productions/Performances
    o E.g., The Unknown Soldier & His Wife, Romeo & Juliet, etc.
  ❑ Subject Headings/Terms
    o E.g., Theater—History, Costume Design, etc.
  ❑ Bibliographic References
    o E.g., Figaro, Gaulois, Journal des Debats, etc.
### Person URI’s Found through Manual Process

<table>
<thead>
<tr>
<th>Total persons identified in Motley metadata = 984</th>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Links have been found for 624 names</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>having Wikipedia / DBPedia links</td>
</tr>
<tr>
<td>having VIAF links</td>
</tr>
<tr>
<td>found by searching viaf.org directly</td>
</tr>
<tr>
<td>found by searching LC Name Authority File</td>
</tr>
<tr>
<td>found by searching WorldCat Identities</td>
</tr>
<tr>
<td>*combined with automatic results</td>
</tr>
<tr>
<td>having Theatricalia links</td>
</tr>
<tr>
<td>having IMDb links</td>
</tr>
<tr>
<td>having IBDb links</td>
</tr>
<tr>
<td>having more than 1 link</td>
</tr>
</tbody>
</table>

*VIAF links for 476 persons (364 not found by manual search) were found using VIAF Auto Suggest

**Represents some overlapping results
# Theater and Play/Performance URI’s Found through Manual Process

**Total theaters identified in Motley metadata = 59**

<table>
<thead>
<tr>
<th>Links were found for 52 theaters</th>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>having Wikipedia / DBPedia links</td>
<td>49 (83%)</td>
</tr>
<tr>
<td>having VIAF links</td>
<td>45 (76%)</td>
</tr>
<tr>
<td>having home page links</td>
<td>36 (61%)</td>
</tr>
<tr>
<td>having other links</td>
<td>16 (27%)</td>
</tr>
<tr>
<td>having more than 1 link</td>
<td>47 (80%)</td>
</tr>
</tbody>
</table>

**Total plays / performances identified in Motley metadata = 127**

<table>
<thead>
<tr>
<th>Links were found for 105 plays / performances</th>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>having Wikipedia / DBPedia links</td>
<td>95 (75%)</td>
</tr>
<tr>
<td>having Theatricalia links</td>
<td>45 (35%)</td>
</tr>
<tr>
<td>having other links</td>
<td>10 (8%)</td>
</tr>
<tr>
<td>having more than 1 link</td>
<td>44 (35%)</td>
</tr>
</tbody>
</table>
Kolb-Proust Archive Entities

<table>
<thead>
<tr>
<th>Total number of names found in the Kolb-Proust dataset = 5,727</th>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Links were found for 1,953 people</td>
<td></td>
</tr>
<tr>
<td>having VIAF links</td>
<td>1,678 (29%)</td>
</tr>
<tr>
<td>having French Wikipedia links</td>
<td>1,236 (22%)</td>
</tr>
<tr>
<td>having English Wikipedia links</td>
<td>999 (17%)</td>
</tr>
<tr>
<td>having other links</td>
<td>264 (5%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total number of notecards in the Kolb-Proust dataset = 8,716</th>
<th>Count of URIs Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations found on notecards</td>
<td>13,923 (~1.6 citations/card)</td>
</tr>
<tr>
<td>Links founds for citations</td>
<td>4,812 (35%)</td>
</tr>
</tbody>
</table>
Observations

• Additional name authority sources needed for special collections
  - Many current sources are focused on authors

• When searching manually for entity links we found:
  - Easiest to start in WorldCat Identities; Google Web Search next best
  - Googling with full names and birth dates usually insufficient, needed to include additional keyword for best results

• Manual clean up of metadata and manual search helps recall:
  - Different name spellings/maiden names/nicknames, slightly different birth/death dates, and looking for contextual clues
What’s the Benefit of All This?
How Do We Use It?
Context Enhancement

PROPERTIES OF THE SKETCH
A schema:VisualArtwork

PERFORMANCE
AND VENUE ENTITIES
REFERENCED

PERSONS
REFERENCED
Preliminary Findings & Conclusions

• Care needs to be taken when mapping legacy metadata to LOD-compliant vocabularies
  - May need to extend with additional entities and properties
  - However can sometimes be rewarded with additional linking properties (e.g., schema:mentions and schema:citation)

• User experiences enriched by adding contextual information
  - Through dynamically added sidebars and clickable links
    - Leverage existing Semantic Web sources
    - Provide opportunities for users to escape the siloed environments of traditional digital libraries
  - However, it is resource-intensive to manually add links, etc. to legacy metadata

• Additional Opportunities for leveraging the Semantic Web remain to be explored
Works in Progress
Pushing Information Back Out to the Semantic Web
Data Visualization
Knowledge Cards on Search Results Pages
Additional Opportunities for leveraging the Semantic Web remain to be explored!

Myung-Ja K. Han (mhan3@Illinois.edu)
Deren Kudeki (dkudeki@illinois.edu)
Timothy W. Cole (t-cole3@Illinois.edu)
Jacob Jett (jjett2@Illinois.edu)