Digital Scriptorium 2.0: 
Toward a community-driven LOD knowledge base and national union catalog for premodern manuscripts

L.P. Coladangelo, College of Communication and Information, Kent State University
Lynn Ransom, Schoenberg Institute for Manuscript Studies, Penn Libraries
Doug Emery, Penn Libraries

Semantic Web in Libraries 2022 (SWIB22)
December 2, 2022
Agenda

Background
- Manuscript Metadata
- Previous Database

Overview of DS 2.0 Data Model
Overview of DS 2.0 Workflows

Decision-making toward Implementation
- Semantic Enrichment

Next Steps
Digital Scriptorium and the Need for Redevelopment
Manuscript Description

**Manuscripts are unique objects** (Bair & Steuer, 2013)  
Requires description of  
content  
physical properties  
context around creation and provenance

**Knowledge about manuscripts can be disparate and debated**  
Requires extensive scholarship  
Revisions in descriptions occur as more is learned

**Imperfect reliance on bibliographic standards**  
No uniformly agreed upon standards for manuscript descriptions  
Barriers to institutions without expertise and resources to implement bibliographic description standards
The “Old” Digital Scriptorium

Institutional Records (any number of formats)

manual input

METS-based spreadsheets

upload to

Digital Scriptorium Database (METS data)
The “Old” Digital Scriptorium

- maintenance of two sets of metadata records
- labor-intensive, error-prone
- limited browsing and searching
- data not optimized for linked data applications
Principles for the DS 2.0 Redevelopment

**Minimal data entry standards**
manuscript location in an institution is sufficient

**Members manage their own metadata (formats and values)**
DS 2.0 uses what is provided and does not correct or add to member metadata

**DS 2.0 does not host images**
primary goal is initial discovery and metadata aggregation but IIIF functionality is possible with provided URIs for manifests

**DS 2.0 will provide semantic enrichment**
metadata is enhanced, where possible, by reconciling values with external authorities as well as linking to internal Name and Manuscript ID Authorities

**Open access**
DS 2.0 data will be open and available for reuse
DS 2.0 Data Model

Backbone of main classes of entities (Koho et al., 2022)
- **Manuscript** (physical manuscript object)
- **DS 2.0 Record** (metadata description of a Manuscript)
- **Holding** (metadata description of holding information about a Manuscript)

Separates the physical object from its catalog record / metadata description
- Conceptualizes the institutional metadata record as a cataloger’s observation of a manuscript object
- Can be revised / overwritten as changes in description and scholarship warrant
- Physical object retains a unique and persistent DS ID despite description and holdings changes
DS 2.0 Data Model as implemented
DS 2.0 Workflow

Members’ structured data (MARC/MARC-XML, METS, TEI, CSV, etc.)

OpenRefine

Agnostic Transition Spreadsheet

DS converts

Data values extracted by metadata element

DS transforms

Semantically enriched metadata recombined

DS gives

DS 2.0
Commentary on the Pauline Epistles (DS1878) (Q3522)

Manuscript metadata collected by UPenn from University of Kansas (MS 9/1:B10)

**In more languages**

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Commentary on the Pauline Epistles (DS1878)</td>
<td>Manuscript metadata collected by UPenn from University of Kansas (MS 9/1:B10)</td>
</tr>
</tbody>
</table>

**Statements**

- **Language as recorded**
  - Latin
    - Language in authority file: Latin

- **Physical description**
  - Extent: f. 1r-v; 325 x 237 mm.
Peter Lombard (Q315347)

medieval priest and theologian
Master of Sentences

In more languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Peter Lombard</td>
<td>medieval priest and theologian</td>
</tr>
<tr>
<td>Spanish</td>
<td>Pedro Lombardo</td>
<td>No description defined</td>
</tr>
<tr>
<td>Traditional Chinese</td>
<td>No label defined</td>
<td>No description defined</td>
</tr>
<tr>
<td>Chinese</td>
<td>伦巴第人彼得</td>
<td>伦巴第人彼得</td>
</tr>
</tbody>
</table>

Statements

<table>
<thead>
<tr>
<th>Statements</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance of</td>
<td>human</td>
<td></td>
</tr>
</tbody>
</table>

0 references

No description defined

Wikidata QID | Q315347

DS ID | DSNME251

0 references

No description defined
LOD Vocabularies for Semantic Enrichment

Centuries
- Getty Art and Architecture Thesaurus (AAT)

Languages
- Wikidata

Materials
- AAT (controlled list)

People, Organizations
- Wikidata

Places
- Getty Thesaurus of Geographic Names (TGN)

Roles
- Controlled list

Terms (Subject and Genre)
- AAT
- OCLC Faceted Application of Subject Terminology (FAST)
- Ligatus Language of Binding Terms (LoBT)
Guiding Choices for Semantic Enrichment

Consider only vocabularies published as LOD
Leaving values unreconciled if an appropriate match cannot be found

Match only where appropriate as indicated by the data
  e.g., MARC subfields for the source for genre terms

Use vocabularies known to GLAM communities
  e.g., sources for places were not identified, but TGN was used

Leverage the value and flexibility of linked data
  e.g., LC derived-strings and the choice to use FAST; Wikidata for languages and their ISO codes

Address limitations in existing authorities
  e.g., use of Wikidata instead of LCNAF or VIAF

Implement in a manageable data processing workflow
  Using reconciliation services available through OpenRefine
Open Documentation

https://github.com/DigitalScriptorium
<table>
<thead>
<tr>
<th></th>
<th>Country</th>
<th>Place</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>France</td>
<td>France</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>49</td>
<td>France</td>
<td>France</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>50</td>
<td>France (Normandy)</td>
<td>France/Normandy</td>
<td><a href="http://vocab.getty.edu/tgn/7002886">http://vocab.getty.edu/tgn/7002886</a></td>
</tr>
<tr>
<td>51</td>
<td>France (Paris)</td>
<td>France/Paris</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>52</td>
<td>France or Flanders</td>
<td>France/Flanders</td>
<td><a href="http://vocab.getty.edu/tgn/7004907">http://vocab.getty.edu/tgn/7004907</a></td>
</tr>
<tr>
<td>53</td>
<td>France, Anjou</td>
<td>France/Anjou</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>54</td>
<td>France, Brittany</td>
<td>France/Brittany</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>55</td>
<td>France, Dauphiné</td>
<td>France/Dauphiné</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>56</td>
<td>France, Franche Comté</td>
<td>France/Franche-Comté</td>
<td><a href="http://vocab.getty.edu/tgn/7007864">http://vocab.getty.edu/tgn/7007864</a></td>
</tr>
<tr>
<td>57</td>
<td>France, Normandy</td>
<td>France/Normandy</td>
<td><a href="http://vocab.getty.edu/tgn/7002886">http://vocab.getty.edu/tgn/7002886</a></td>
</tr>
<tr>
<td>58</td>
<td>France, Northern</td>
<td>Northern France</td>
<td><a href="http://vocab.getty.edu/tgn/4004036">http://vocab.getty.edu/tgn/4004036</a></td>
</tr>
<tr>
<td>59</td>
<td>France, Picardy</td>
<td>France/Picardy</td>
<td><a href="http://vocab.getty.edu/tgn/7002890">http://vocab.getty.edu/tgn/7002890</a></td>
</tr>
<tr>
<td>60</td>
<td>France, probably between 1425 and 1450</td>
<td>France</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>61</td>
<td>France?</td>
<td>France</td>
<td><a href="http://vocab.getty.edu/tgn/1000070">http://vocab.getty.edu/tgn/1000070</a></td>
</tr>
<tr>
<td>62</td>
<td>France, Netherlands</td>
<td>France/Netherlands</td>
<td><a href="http://vocab.getty.edu/tgn/7001884">http://vocab.getty.edu/tgn/7001884</a></td>
</tr>
<tr>
<td>63</td>
<td>France, Champagne</td>
<td>France/Champagne</td>
<td><a href="http://vocab.getty.edu/tgn/7603890">http://vocab.getty.edu/tgn/7603890</a></td>
</tr>
<tr>
<td>64</td>
<td>Germany</td>
<td>Germany</td>
<td><a href="http://vocab.getty.edu/tgn/7000084">http://vocab.getty.edu/tgn/7000084</a></td>
</tr>
<tr>
<td>65</td>
<td>Germany</td>
<td>Germany</td>
<td><a href="http://vocab.getty.edu/tgn/7000084">http://vocab.getty.edu/tgn/7000084</a></td>
</tr>
<tr>
<td>66</td>
<td>Germany (Nuremberg)</td>
<td>Germany/Nuremberg</td>
<td><a href="http://vocab.getty.edu/tgn/7000084">http://vocab.getty.edu/tgn/7000084</a></td>
</tr>
<tr>
<td>67</td>
<td>Germany, Bavaria</td>
<td>Germany/Bavaria</td>
<td><a href="http://vocab.getty.edu/tgn/7003699">http://vocab.getty.edu/tgn/7003699</a></td>
</tr>
<tr>
<td>68</td>
<td>Germany, Bavaria or west Austria</td>
<td>Germany/Bavaria/Austria</td>
<td><a href="http://vocab.getty.edu/tgn/7003699">http://vocab.getty.edu/tgn/7003699</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>genre_ac_recorded</th>
<th>vocabulary</th>
<th>source_authority_url</th>
<th>authorized_label</th>
<th>structured_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Abridgements</td>
<td>alt</td>
<td></td>
<td>abridgements</td>
<td><a href="http://vocabs.getty.edu/alt/300023489">http://vocabs.getty.edu/alt/300023489</a></td>
</tr>
<tr>
<td>2</td>
<td>Accounts</td>
<td>alt</td>
<td></td>
<td>accounts</td>
<td><a href="http://vocabs.getty.edu/alt/3000145802">http://vocabs.getty.edu/alt/3000145802</a></td>
</tr>
<tr>
<td>3</td>
<td>Amulets</td>
<td>alt</td>
<td></td>
<td>amulets</td>
<td><a href="http://vocabs.getty.edu/alt/3000065803">http://vocabs.getty.edu/alt/3000065803</a></td>
</tr>
<tr>
<td>4</td>
<td>Annals and chronicles</td>
<td>lcgt</td>
<td></td>
<td>Annals and chronicles</td>
<td><a href="http://dwc.worldcat.org/fast/19190982">http://dwc.worldcat.org/fast/19190982</a></td>
</tr>
<tr>
<td>5</td>
<td>Annotations</td>
<td>alt</td>
<td></td>
<td>annotations</td>
<td><a href="http://vocabs.getty.edu/alt/300021600">http://vocabs.getty.edu/alt/300021600</a></td>
</tr>
<tr>
<td>6</td>
<td>Anthologies</td>
<td>alt</td>
<td></td>
<td>anthologies</td>
<td><a href="http://vocabs.getty.edu/alt/3000060377">http://vocabs.getty.edu/alt/3000060377</a></td>
</tr>
<tr>
<td>7</td>
<td>Art</td>
<td>alt</td>
<td></td>
<td>Art</td>
<td><a href="http://dwc.worldcat.org/fast/8151777">http://dwc.worldcat.org/fast/8151777</a></td>
</tr>
<tr>
<td>8</td>
<td>Artificial grains</td>
<td>lbbt</td>
<td></td>
<td>artificial grains</td>
<td><a href="http://viaf.org/ltb/concept/1507">http://viaf.org/ltb/concept/1507</a></td>
</tr>
<tr>
<td>9</td>
<td>Blind tooled (decoration)</td>
<td>lbbt</td>
<td></td>
<td>blind tooled (decoration)</td>
<td><a href="http://viaf.org/ltb/concept/2288">http://viaf.org/ltb/concept/2288</a></td>
</tr>
<tr>
<td>10</td>
<td>Books of hours</td>
<td>alt</td>
<td></td>
<td>books of hours</td>
<td><a href="http://vocabs.getty.edu/alt/3000254577">http://vocabs.getty.edu/alt/3000254577</a></td>
</tr>
<tr>
<td>11</td>
<td>Breviaries</td>
<td>lbbt</td>
<td></td>
<td>breviaries</td>
<td><a href="http://viaf.org/ltb/concept/12813">http://viaf.org/ltb/concept/12813</a></td>
</tr>
<tr>
<td>12</td>
<td>Calendars (documents by form)</td>
<td>lbbt</td>
<td></td>
<td>calendars (documents)</td>
<td><a href="http://vocabs.getty.edu/alt/3000267411">http://vocabs.getty.edu/alt/3000267411</a></td>
</tr>
<tr>
<td>13</td>
<td>Choirbooks</td>
<td>alt</td>
<td></td>
<td>choirbooks</td>
<td><a href="http://vocabs.getty.edu/alt/300026422">http://vocabs.getty.edu/alt/300026422</a></td>
</tr>
<tr>
<td>14</td>
<td>Chronicles</td>
<td>alt</td>
<td></td>
<td>chronicles</td>
<td><a href="http://vocabs.getty.edu/alt/3000253611">http://vocabs.getty.edu/alt/3000253611</a></td>
</tr>
<tr>
<td>15</td>
<td>Clasps (components)</td>
<td>lbbt</td>
<td></td>
<td>clasps (components)</td>
<td><a href="http://viaf.org/ltb/concept/1254">http://viaf.org/ltb/concept/1254</a></td>
</tr>
<tr>
<td>16</td>
<td>Codices</td>
<td>alt</td>
<td></td>
<td>codices</td>
<td><a href="http://vocabs.getty.edu/alt/300022400">http://vocabs.getty.edu/alt/300022400</a></td>
</tr>
<tr>
<td>17</td>
<td>Codices (bound manuscripts)</td>
<td>alt</td>
<td></td>
<td>codices (bound manuscripts)</td>
<td><a href="http://vocabs.getty.edu/alt/300022400">http://vocabs.getty.edu/alt/300022400</a></td>
</tr>
<tr>
<td>18</td>
<td>Commentaries</td>
<td>alt</td>
<td></td>
<td>Commentaries</td>
<td><a href="http://dwc.worldcat.org/fast/1437323">http://dwc.worldcat.org/fast/1437323</a></td>
</tr>
<tr>
<td>19</td>
<td>Commentaries</td>
<td>alt</td>
<td></td>
<td>commentaries</td>
<td><a href="http://vocabs.getty.edu/alt/3000236966">http://vocabs.getty.edu/alt/3000236966</a></td>
</tr>
<tr>
<td>20</td>
<td>Commentaries</td>
<td>alt</td>
<td></td>
<td>commentaries</td>
<td><a href="http://vocabs.getty.edu/alt/3000236966">http://vocabs.getty.edu/alt/3000236966</a></td>
</tr>
</tbody>
</table>

string value is qualified by individual authority values, which connect to linked data vocabularies
Next Steps for DS 2.0

Move from Prototype to Beta Version
  Taking in all legacy data (existing METS) and member data (MARC, TEI, CSV, etc.)

Development of a User Interface
  Traditional online catalog interface for existing users

Ongoing analysis of Data Quality
  Continuing to look at correctness and completeness of metadata

Providing Services to Users
  e.g., example SPARQL queries, workshops on Wikibase
Thank you!

https://digital-scriptorium.org/

Twitter: dscriptorium2_0

L.P. Coladangelo
lcoladan@kent.edu
@lpc359