Digital Scriptorium 2.0

A National Union Catalog of Premodern Manuscripts in North American Collections for the Digital Age

2nd Stakeholder Meeting to Present Plans for Data Model and Workflow

April 30, 2021
## Update on Progress

<table>
<thead>
<tr>
<th>Activity</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
<th>Month 6</th>
<th>Month 7</th>
<th>Month 8</th>
<th>Month 9</th>
<th>Month 10</th>
<th>Month 11</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hire Project Manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membership Surveys: 1) Member Needs and Expectations and 2) Value Proposition Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholder Meetings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Data Content Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental scan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Development &amp; Implementation Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Reports to DS Board of Directors and Membership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Design and Implementation Plan Finalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Original Workplan
## Revised Work Plan

<table>
<thead>
<tr>
<th>Activity</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
<th>Month 5</th>
<th>Month 6</th>
<th>Month 7</th>
<th>Month 8</th>
<th>Month 9</th>
<th>Month 10</th>
<th>Month 11</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hire Project Manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membership Surveys: 1) Member Needs and Expectations and 2) Value Proposition Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholder Meetings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Data Content Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental scan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Development &amp; Implementation Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Reports to DS Board of Directors and Membership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Model Design and Implementation Plan Finalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Agenda

1pm-2pm

- Overview of Vision for DS 2.0 (LR)
  - Discussion

- Data model explanation + Wikibase intro (ET)
  - Discussion

2pm-3pm

- Data contribution workflow explanation (ET)
  - Discussion

- Implementation plan, timeline, and plan for further funding (LR)
  - Discussion
DS 2.0 Planning Project Acknowledgements

Project Team
- Emma Thomson, Project Manager
- Lynn Ransom, Project Director
- Doug Emery, Technical Advisor
- Mikko Koho, Technical Consultant, University of Helsinki
Mapping Manuscript Migrations (MMM) is a semantic portal for finding and studying pre-modern manuscripts and their movements, based on linked collections of the Schoenberg Institute for Manuscript Studies, the Bodleian Libraries, and the Institut de recherche et d’histoire des textes.

Select a perspective to search and browse the MMM data:

- **Manuscripts**
  - Physical manuscript objects
- **Works**
  - Intellectual content of manuscripts
- **Events**
  - Events related to manuscripts
- **Actors**
  - People and institutions related to manuscripts
- **Places**
  - Places related to manuscripts

Images used under licenses from Shutterstock.com, Wikimedia, and Others.
DS 2.0 Planning Project Acknowledgements

Project Team
- Emma Thomson, Project Manager
- Lynn Ransom, Project Director
- Doug Emery, Technical Advisor
- Mikko Koho, Technical Consultant, University of Helsinki

Steering Committee
- Debra Taylor Cashion, Saint Louis University
- Ray Clemens, Beinecke Library, Yale University
- Lisa Fagin Davis, Medieval Academy of America
- Tamar Evangelestia-Dougherty, Cornell University
- Elizabeth Hebbard, Indiana University-Bloomington
DS 2.0 Planning Project Acknowledgements

Project Team
- Emma Thomson, Project Manager
- Lynn Ransom, Project Director
- Doug Emery, Technical Advisor
- Mikko Koho, Technical Consultant, University of Helsinki

Steering Committee
- Debra Taylor Cashion, Saint Louis University
- Ray Clemens, Beinecke Library, Yale University
- Lisa Fagin Davis, Medieval Academy of America
- Tamar Evangelestia-Dougherty, Cornell University
- Elizabeth Hebbard, Indiana University-Bloomington

Penn Libraries Consultants
- Jim Hahn, Head of Metadata Research
- Amey Hutchins, Manuscripts Cataloging Librarian
- John Mark Ockerbloom, Digital Library Strategist & Metadata Architect
- Kelly Tuttle, formerly Project Cataloger, Manuscripts of the Muslim World
- Sean Quimby, Associate University Librarian & Director of the Jay I. Kislak Center for Special Collections, Rare Books and Manuscripts
DS 2.0 Planning Project Acknowledgements

Board of Directors
- Debra Taylor Cashion (President; 2015-2021)
- Janine Pollock (Vice President; 2015-2021)
- Lynn Ransom (Secretary; 2015-2021)
- E. C. Schroeder (Treasurer; 2015-2021)
- Lynne Grigsby (Technology Host; 2009-present)
- David Faulds (Director at Large; 2018-2021)
- Vanessa Wilkie (Director at Large; 2015-2021)
- Cherry Williams (Director at Large; 2015-2021)

Advisory Council
- Raymond Clemens, Yale University
- Lisa Fagin Davis, Medieval Academy of America; Simmons College (Chair)
- Consuelo Dutschke, Emerita, Columbia University
- Barbara Shailor, Yale University
- William P. Stoneman, Emeritus, Harvard University
DS 2.0 Objectives
Established at the Beinecke Library DS 2.0 Planning Meeting
(February 2019)

● Be a member-supported National Union Catalog for all premodern manuscripts in North America

● Provide a low barrier to both the contribution and use of metadata and images for premodern manuscripts from all manuscript cultures

● Build upon a clear, adaptable, and scalable data model

● Employ linked and linkable data

● Support interoperability with other manuscript research projects at regional, national, and international levels

● Enable content contributors to maintain management and ownership of data while sharing in the continuing benefits of national collaboration
DS 2.0 Principles

- As a national union catalog, DS 2.0’s primary function will be to enable researchers to find premodern manuscripts in US collections, including non-European manuscripts.

- DS 2.0 will require minimal standards for data entry. Metadata only identifying a manuscript’s location in an institution will be sufficient.

- Members will manage their own manuscript metadata in their institutional formats. DS 2.0 will use what members provide and will not correct or add to a member’s metadata.

- DS 2.0 will not host images, but will provide IIIF functionality to view images in platform. DS 2.0 will work with members to find hosting solutions as needed.

- DS 2.0 will enhances metadata with Name, Place, and in-house Manuscript ID Authorities and will make DS 2.0 data available for reuse.
Vision for DS 2.0

Members' structured data
Vision for DS 2.0

DS converts

Agnostic Transition Spreadsheet

Members' structured data
Vision for DS 2.0

Members’ structured data

DS converts

Agnostic Transition Spreadsheet

DS transforms

DS 2.0
Vision for DS 2.0

DS converts Members’ structured data to Agnostic Transition Spreadsheet, which DS transforms into DS 2.0. DS gives DS 2.0 to the world.
**DS 2.0 Proposed Workflow**

**Member Responsibility**

- Provide best available data in structured format, e.g.:
  - DS files
  - MARC
  - TEI
  - EAD
  - SDBM csv
  - ?

**DS 2.0 responsibilities**

- Convert CSV into RDF
- Assign and manage MS ID Authority
- Data reconciliation for Names, Places
- Manage LOD-based DS Authorities for Names and Places, linking and contributing to external authorities
- Supply links back to member records and images
- Build and maintain discovery interface with keyword and faceted searching
- Maintain and regularly update DS 2.0 data
- Make DS data available for re-use as LOD via SPARQL endpoint
- Provide IIIF functionality via Mirador or similar software for image viewing
- Assist data transfers between holding members as needed; in case of no transfer, maintain last holding record
DS 2.0 Outcomes

- A member supported national union catalogue of manuscripts in US collections
- Management of MS Authority Records and Identifiers for US manuscripts
- Partner organization for member digitization and cataloging projects
- Access point for national and international partnerships involving US manuscript data
- Build a program to train catalogers and students in best practices for manuscript description
DS 2.0 Data Model Principles

• DS 2.0’s primary function: enable researchers to find premodern manuscripts in US collections

• Minimal standards for data entry

• Members manage their own metadata in their own formats. We use what we get and do not correct or add to it even if metadata profile is “incomplete”

• DS 2.0 enhances metadata through Name, Place, and in-house DS ID authorities and makes all data available for reuse

• Unique manuscript number (DS ID) created for each US manuscript at the time it is first entered into DS 2.0
DS 2.0 Data Fields

• DS 2.0 information
  • DS ID (unique, permanent identifier for each manuscript)
  • Date added
  • Date last updated

• Holding information
  • Holding institution
  • Holding institution ID number (shelfmark)
  • Link to holding institution’s record

• Descriptive information
  • Production place
  • Production date (including Dated MS)
  • Title
  • Genre
  • Subject
  • Author
  • Artist
  • Scribe
  • Language
  • Material
  • Physical Description
  • Former owners
  • General notes
  • Acknowledgments
Linked Open Data Basics

- Structured data that is published online and links to other data on the Web
- Designed to be processed by computers
  - Data constructed using the RDF conceptual model
  - Uses URIs (uniform resource identifiers) to create unique and unambiguous references to identify information resources
- LOD can turn the internet into one giant database (the Semantic Web)
RDF Basics: the triple
RDF Basics: a triple example for DS 2.0

[Diagram showing a triple example with subjects and predicates]

- **DS 2.0 Record**
- **hasTitle**
- **De civitate Dei**
RDF Basics: a triple example for DS 2.0

Technology for DS 2.0: Wikibase

• Open-source software for creating LOD

• Used by many other similar projects
  • Biblissima
  • GND authority files

• Data exports in multiple formats (JSON, RDF/XML, N3, YAML, etc.), along with built in SPARQL endpoint

• Supports data reconciliation via OpenRefine
DS 2.0 Contribution Workflow Objectives

• Provide a low barrier to both the contribution and use of metadata and images for premodern manuscripts from all manuscript cultures

• Enable content contributors to maintain management and ownership of data while sharing in the continuing benefits of national collaboration
DS 2.0 Member Responsibilities

• Maintain membership in Digital Scriptorium

• Share your manuscript holdings data with DS 2.0 on a regular basis (cycle still to-be-determined)
DS 2.0 Proposed Workflow

Member Responsibility

- Provide best available data in structured format, e.g.:
  - DS files
  - MARC
  - TEI
  - EAD
  - SDBM csv
  - ?

DS 2.0 responsibilities

- Convert CSV into RDF
- Assign and manage MS ID Authority
- Data reconciliation for Names, Places
- Manage LOD-based DS Authorities for Names and Places, linking and contributing to external authorities
- Supply links back to member records and images
- Build and maintain discovery interface with keyword and faceted searching
- Maintain and regularly update DS 2.0 data
- Make DS data available for re-use as LOD via SPARQL endpoint
- Provide IIIF functionality via Mirador or similar software for image viewing
- Assist data transfers between holding members as needed; in case of no transfer, maintain last holding record
Additional DS 2.0 Staff Responsibilities

• Collaborate with national and international partners
• Assist with cataloging and digitization as needed (through grant funding or other routes)
• Assign Dated value to dated manuscripts for http://palaeographia.org/cipl/cipl.htm
• Support maintenance of MSID if a manuscript leaves a DS member institution
Implementation Timeline

- Planning Year (ending August 31, 2021)
- Year 1: The “Bridge Year”
- Years 2-4: Implementation
Implementation Timeline

- Planning Year (ending August 31, 2021)
  - Prototype built and trialed
  - Results of trial data entry reviewed
  - Present results to membership (July)

- Year 1: The “Bridge Year”

- Years 2-4: Full Implementation
Implementation Timeline

● Planning Year (ending July 31, 2021)

● Year 1: The “Bridge Year” (starting August 1, 2021)
  ○ Begin beta testing prototype
  ○ Import Penn & Beinecke Library mss data and assess results
  ○ Crosswalk of DS 1.0 data into DS 2.0 environment
  ○ Establish data ingest workflow from spreadsheet
  ○ Perform Name and Place authority reconciliation in test data
  ○ Develop MS ID management workflows
  ○ Submit IMLS National Leadership Grant for Libraries
    1. Likely timeline: pre-proposal due October 2021; accepted proposal due January 2022; notifications in July 2022; grant begins August/September 2022). Funding up to $1 million, with matching cost-share

2. Years 2-4: Implementation
Implementation Timeline

- Planning Year (ending July 31, 2021)
- Year 1: The “Bridge Year” (starting August 1, 2021)
- Years 2-4: Implementation
Implementation Timeline

- Planning Year (ending July 31, 2021)
- Year 1: The “Bridge Year” (starting August 1, 2021)
- Years 2-4: Implementation
  - Assuming we get funded ...
Funded Implementation Plan

- Years 2-4: Implementation
  - Year 2: Complete beta testing, implement changes, add more test data, strengthen and build membership
  - Year 3:
    - Transfer data from members' institutional catalogs
    - Assist members in creating or updating manuscript data through postdoctoral and graduate student fellowship program
    - Develop manuscript description training program, with possible funding from the NEH Preservation and Access Training Grant Program
  - Year 4:
    - Continue adding member data
    - Implement NEH-funded manuscript description training program
    - Review membership requirements and roles to support DS 2.0 beyond implementation
Funded Implementation Plan

- Years 2-4: Implementation
  - Year 2: Complete beta testing, implement changes, add more test data
  - Year 3:
    - Transfer data from members’ institutional catalogs
    - Assist members in creating or updating manuscript data through postdoctoral and graduate student fellowship program
    - Develop manuscript description training program, with possible funding from the NEH Preservation and Access Training Grant Program
  - Year 4:
    - Continue adding member data
    - Implement NEH-funded manuscript description training program
    - Review membership requirements and roles to support DS 2.0 beyond implementation

But if we don't get funded ...
Non-Funded Implementation Plan

We believe that we could at least (thanks to Bridge Year funding):

- Complete beta testing, implement changes, add more test data...but at a slower pace
- Transfer data from members’ institutional catalogs
- Assist members in creating or updating manuscript data through postdoctoral and graduate student fellowship program
- Develop manuscript description training program, with possible funding from the NEH Preservation and Access Training Grant Program
- Continue adding member data
- Implement manuscript description training program
- Review membership requirements and roles to support DS 2.0 beyond implementation
Post-Implementation Plans

- Continued management of DS 2.0 at a host institution with a dedicated full-time Project Manager supported by membership fees
- Continued growing membership
- Build education programming for best practices in manuscript description in a digital age
  - NEH Preservation and Access Training Grant