Digital Scriptorium 2.0 Project Report III
June 10, 2021

Introduction

In July 2020, the Institute of Museum and Library Services (IMLS) awarded Penn Libraries a $100,000 National Leadership for Libraries Grant to oversee a one-year planning period for the future of Digital Scriptorium (DS), a consortium of US institutions who are dedicated to providing open access to their manuscript data and images. Since 1997, DS has hosted an online platform and database to facilitate this mission, but its purpose and technical infrastructure needs redevelopment in order to remain viable in the 21st century. The project team has been working throughout the grant period to refine the purpose and scope of the DS platform in order to ensure its technical and financial sustainability in the years to come. DS's ultimate goal is to become the national union catalogue of pre-modern manuscripts held in US collections.

The DS 2.0 project team is pleased to report that they have completed all the objectives of the planning year, including the development of the DS 2.0 data model, a new data contribution workflow, and plans for implementation and further funding. These proposals were presented to DS members and other stakeholders on April 30, 2021. 57 attendees from over 35 institutions signed up for the virtual event. Participants discussed the plans with the project team and provided positive feedback for the implementation plans both during the meeting and in their responses to an exit survey distributed afterwards. A summary of the major points discussed at the meeting appear below.

Principles of DS 2.0

Over the course of the planning year, the project team honed the principles that will guide the production of DS 2.0. The core goal of the project is to enable researchers to find premodern manuscripts in US collections, including manuscripts produced outside of Europe. Minimal standards for data entry will lower the barrier for contribution, allowing institutions to submit as much or as little data as they are able. The only information required will be data identifying a manuscript's location in an institution. DS members will manage their own data in their own local formats. DS 2.0 will use what members provide without correcting or adding to it. DS 2.0 will not host images but will provide IIIF functionality within the platform, and DS staff will help members who need image hosting assistance find solutions as needed. DS 2.0 will enhance metadata with authority files for names, places, and manuscripts, and make this data available for reuse as linked open data.

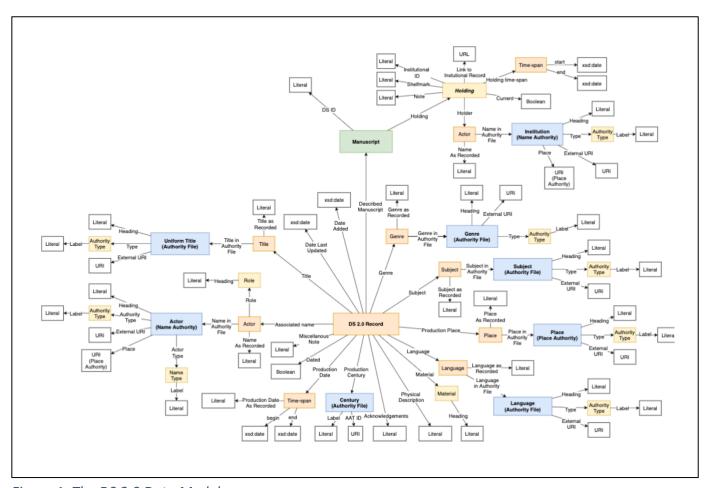## DS 2.0 Data Model and Contribution Workflow

Figure 1: The DS 2.0 Data Model

At the center of the new data model is the DS 2.0 Record, where all of the descriptive information about a manuscript is stored. This is the data that DS members will contribute based on their own institutional records. Data points include, but are not limited to, associated names, titles, subjects, production places, languages, materials, physical description, production dates, and a generic note field. This data will be refreshed regularly and will change when institutions update their local catalog records.

Moving up on the diagram from the DS 2.0 Record is the Manuscript data. This is administrative metadata about a manuscript's existence in DS, rather than descriptive information about the physical manuscript. The DS ID is a unique identifier that will be assigned to each manuscript upon the creation of its record in DS 2.0. This ID will never change, even if a manuscript changes owners. The Holding data relates to the DS member who contributes information about this manuscript. In addition to containing an authorized name for that holding institution, the

Holding data includes a link to the institution's local record (such as a MARC record), as well as institutional ID numbers, shelf marks, and a general note field. DS 2.0 will0 also store the start and end date that an institution contributed data to DS, and state whether the institution is a current or former holder. When a manuscript changes owners, previous holding information will be preserved.

To contribute to DS 2.0, an institution must be a DS member or associate. Institutions will provide data about their manuscripts in a structured format, including but not limited to MARC, TEI, EAD, or CSV files. Records from the current version of DS will also be accepted. DS staff will take over the rest of the data contribution process. Staff will convert the institutional data against the DS 2.0 data model, using a transition spreadsheet which will then be uploaded into the DS 2.0 database. Authority control will be applied to appropriate data fields and managed within DS 2.0. All data will be made available online as linked open data, accessible at any time via a SPARQL endpoint.

Plans for Implementation and Funding

In October 2021, the project team will submit a preproposal to the IMLS National Leadership Grant for Libraries program for implementation funding. If accepted, a full application will be due in March 2022, with notification of award announced in July 2022. Thanks to the contributions of several generous DS members, the project has enough funds to continue employing a full-time project manager for DS 2.0 in the year between the planning grant and anticipated implementation funding. During this bridge year (August 2021-July 2022), the project team will import manuscript data from the University of Pennsylvania and Beinecke Libraries, crosswalk data from the current DS database into DS 2.0, establish the data ingest workflow, and develop authority management for names, places, and manuscript identifiers.

If funded, the three-year full implementation of DS 2.0 would begin in August 2022. The first year of implementation will see the completion of beta testing and data contribution workflow development. In the second year, data transfer from DS members' institutional catalogs will begin. Postdoctoral and graduate student fellowships will be offered to assist DS members in creating or updating their manuscript data. The project team will also develop a manuscript description training program, with possible funding through the NEH Preservation and Access Training Grant Program. The final year of implementation will include the implementation of this training program, along with continued data contributions from members. DS membership requirements and roles will be reviewed to create a sustainable financial plan for DS 2.0 after implementation funding ends.

If DS 2.0 does not receive implementation funding, the project team can still complete beta testing, implement needed changes, and add more test data. Data will still be transferred from members' institutional catalogs into DS 2.0. The manuscript description training program will be developed, with grant support for this program secured via a different funding route. However, without funding all of these goals will be met at a much slower pace. In this scenario, membership requirements and roles will be immediately reviewed and updated to better support DS.

Conclusion and Looking Ahead

For the remaining two months of the planning period, the project team will develop a prototype of the DS 2.0 database in a local Wikibase instance. This will provide an opportunity to test the new data model, produce sample records, and trial the data contribution workflow. The results of this trial period will be presented to DS stakeholders later this summer. The project team would like to extend their gratitude to everyone who contributed to the success of this planning grant, including the IMLS, the University of Pennsylvania, the DS Board of Directors, and all of the DS member institutions and stakeholders.