



Digital Scriptorium 2.0 - Current Data Assessment

February 17, 2021

As the Digital Scriptorium (DS) consortium prepares to redevelop its database and platform, it is important to understand the quality and quantity of the data it's already working with. DS has collected data and images about manuscripts since 1997, resulting in a dataset that currently contains 8543 manuscripts and 76101 images from 36 institutions across the United States. DS's existing database software does not create timestamps for any of this data. No information regarding the initial creation date or the last update for any of the data or images in DS exists within the platform. Understanding the age of the data within DS is critical to evaluating how much of the current DS data must be transformed or cross-walked into the next iteration of the DS database, DS 2.0. If the manuscript descriptions are so old that they should be updated with more recent information anyway, then crosswalking this old data is not the most efficient strategy. The purpose of this data assessment is to analyze the current DS platform's data content and evaluate the state of each institution's data viability in a DS 2.0 environment. We contacted every DS member or associate institution that currently has data in DS and asked them two questions:

1. When was the last time your institution updated its DS data?
2. Is information about your institution's manuscripts available on any other digital platform, such as your institutional catalog or another digital library?

32 out of the 33 current DS member and associate institutions responded to our request for information.¹ We are grateful to all consortium members for their participation in this assessment, which often involved considerable time and effort to locate information about DS work completed many years ago.

¹ Additionally, 3 former members of DS still have data within the platform. These institutions were not included in the data assessment because their information cannot be updated and will not be migrated to the DS 2.0 platform.

In response to our question regarding data updates, 5 institutions stated that they have updated their data within the last 5 years. An additional 8 have definitely updated their data within the last 10 years. 3 institutions could not provide any estimate at all. Put another way, only 40% of DS contributors have certainly updated their information within the last decade. In the course of our correspondence for this report, many institutions commented that the workflow for updating data in DS is very labor-intensive and prevents them from even attempting it. These statistics and comments all highlight the necessity of having a straightforward and regular update process for DS 2.0. DS cannot be a reliable source of information if it neither tracks updates nor offers a sustainable path for performing them.

Our second question asked institutions if they publish their manuscript data or images on any other digital platform, such as an institutional catalog or other digital library. 20 institutions do provide access to their manuscripts on other digital platforms, often via MARC records in their online public access catalogs. This is important as the DS 2.0 contribution workflow will capitalize on institutional records that already exist and simply crosswalk data from those records into DS 2.0. 13 institutions responded that they are dependent on DS as the sole access point for their manuscript data and images. This amounts to 1685 manuscript records (20% of the total manuscripts in DS) and 5497 individual images (7% of the total images in DS). While it will be possible to crosswalk manuscript data from the current DS platform into DS 2.0, the same cannot necessarily be said for images.

The DS 2.0 interface will access manuscript images via the International Image Interoperability Framework (IIIF) technical specifications. IIIF provides a common set of standards and tools that streamline image sharing across organizations and platforms. Institutions that already utilize IIIF to publish their digitized manuscripts will be able to easily share their images with the DS 2.0 interface. However, many DS member institutions have yet to adopt IIIF, including the 13 institutions who rely on the current DS platform as their image host. Since the DS consortium does not have permanent technical support or an institutional home, continuing to serve as an image repository is not sustainable. The DS 2.0 project team is exploring a variety of other options to facilitate a hosting solution for DS members who cannot implement IIIF standards for their own images. If your institution is willing and has the capacity to host IIIF images on behalf of the DS consortium, please contact us.

The time and resources required to crosswalk the necessary legacy DS data into the new platform will be budgeted for in the DS 2.0 implementation plan. We now know just how many manuscript records that includes, and a rough idea of how many images. One component that has yet to be determined is the total number of DS member and associate institutions that cannot host their own IIIF images. We know that number is at least 13, since that's how many use the current DS platform as their sole discovery service (and it does not use IIIF), but we haven't yet asked the DS members who do publish their images on other platforms whether those images are published using IIIF. That question, along with other questions related to the implementation workflow of DS 2.0, will be included in a forthcoming member survey. This

survey will be circulated in advance of the final stakeholder meeting of the DS 2.0 Planning Grant, to occur later this year before the grant wraps up in July 2021.