Minnesota Digital Library and HathiTrust Prototype an Image Preservation Archive

From September through December 2010 the Minnesota Digital Library (MDL), a state-wide coalition of cultural heritage institutions, worked with HathiTrust to add image content from MDL's Minnesota Reflections database and the Minnesota Historical Society (MHS) to the HathiTrust Digital Library as a prototype preservation archive. In addition to exploring a state-wide approach to addressing growing digital preservation needs, the project also intended to help HathiTrust make progress against its long-term objective to support formats beyond books and journals. Nearly 50,000 images from Minnesota Reflections and MHS were ingested into HathiTrust for ingest in January 2011. We will share some of the lessons learned related to policy, standards, and technology by both the Minnesota and HathiTrust teams from this early effort to engage partners in the ingest preparation process and take HathiTrust beyond the book.

While building this preservation archive MDL and HT learned a number of lessons:

- Master images at local institutions are often not formatted as required by HT and require transformation and the addition of embedded metadata. Many of these master images also lack fixity checks. The format requirements of HT may present a high-bar for potential participants.
- Items in the preservation archive require unique identifiers and that identifier namespace needs continual careful attention as new collections are added.
- Mapping metadata from local systems is difficult to routinize and would require ongoing attention in a long-term preservation effort. Properly packaging items for HT can also be time consuming. The different perspectives of MDL and HT concerning metadata has resulted in differences of directional intent over the project period, some of which have been resolved during this pilot project, and others of which must be revisited before a long-term program is undertaken.
- A programmer would be required in any long-term effort to integrate new collections, building scripts to do metadata mapping and packaging of objects.
- The workflow, tight timeline, and the fact that MDL received and had to filter out sensitive data from one partner, highlighted the critical importance of trust in their relationship as they collaborated to complete the archiving task.
- Image data of the sort in Minnesota Reflections is quite a bit larger than the page images of books currently found in HT collections. This creates challenges for data transfer and package ingest.
- Local institutions may be more sensitive about image dissemination than HT expects. Rights issues have posed key challenges for MDL and HT during this pilot project. The project partners currently hold different expectations and requirements regarding rights and display.
- Once descriptions are ingested into HT, only the catalog information can usually be changed.
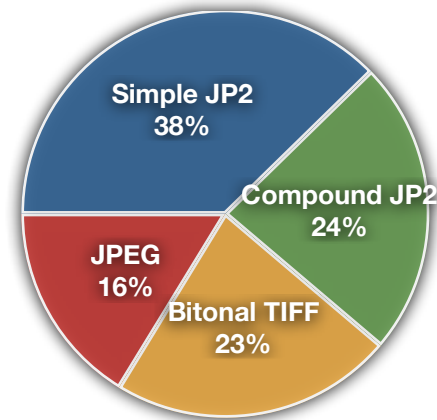- This model would cost about $1 per image up front and $0.10 per image in ongoing maintenance.

Preservation services operate on a continuum from bit-level services to "full" preservation services that include the maintenance of a fully operational access-oriented content catalog. HT is in the highest end of this continuum. In this pilot project, MDL has explored the processes and workflows that would need to be actualized by a wide range of MN institutions in order to participate in HT preservation services.
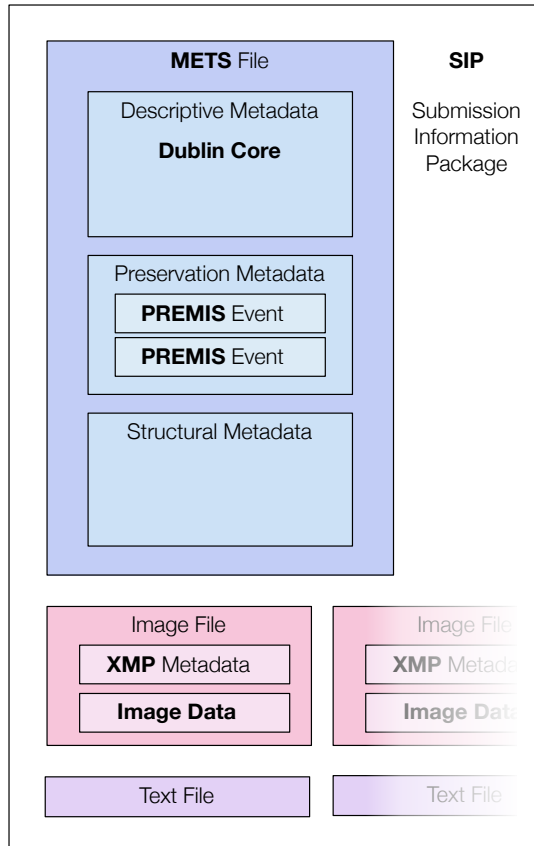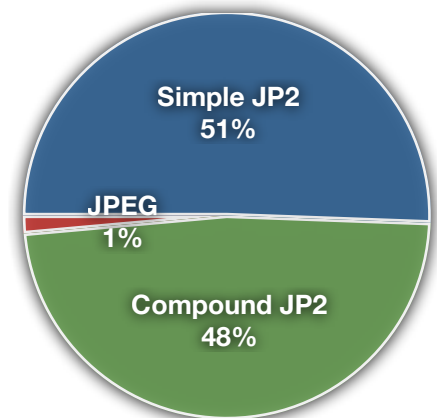

MDL transferred data to HathiTrust using jointly devised SIP. The structure of the SIP is outlined to the left, and the data transferred is described below.

	Items	GB
Simple JP2	22,186	429
Compound JP2	13,844	407
Compound Bitonal TIFF	13,272	1
JPEG	9,575	12
Total	49,302	849

Number of Items Transferred

Our full report is available online:

http://mndigital.org/projects/preservation/

For further information contact:

John Butler, j-butl@umn.edu

Associate University Librarian for Information Technology, University of Minnesota

John Weise, jweise@umich.edu

Head, Digital Library Production Service, University of Michigan

Eric Celeste, efc@clst.org

Consultant, Minnesota Digital Library

Amount of Data Transferred