

Minnesota Digital Library Preservation Options



What we've accomplished with HathiTrust

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. From this experience we learned that there remain a number of questions about the requirements of HathiTrust and the capacity of MDL that might impact a decision of how to best accomplish this kind of digital preservation over the long term.



Some of our questions (referring to the report at <http://www.mndigital.org/projects/preservation>)

What makes a “master” a “master”? When we preserve a digital object, do we need to preserve exactly what we have, or can it be OK to preserve something that can recreate what we have? (See p. 20, *our masters had to be reformatted to JPEG2000 and have metadata embedded.*)

How careful can we be about identifying material? Can we ensure that we provide something as basic as a unique identifier with every digital object we preserve? Can MDL help do this? (See p. 22, *even MDL Reflections found duplicate identifiers, MDL would have to manage a “namespace.”*)

How much metadata do we need in a preservation archive? Digital objects can be difficult to keep track of and some metadata can help us remember what we've saved, but how much is enough? (See p. 26, *culling metadata from a variety of sources is complex, and metadata about preservation events can be thorny.*)

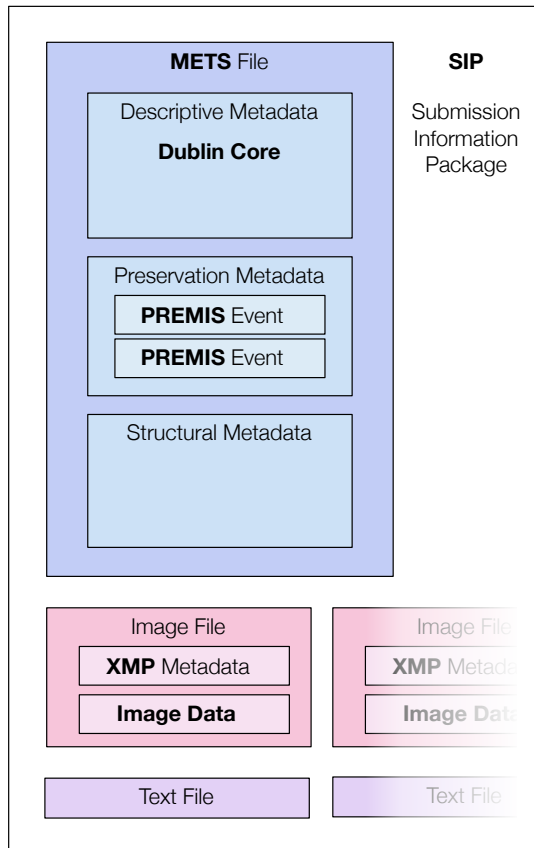
The process with HathiTrust was quite programmer-intensive, can we simplify the requirements or find an option with a lower bar that still does what we need it to do? (See p. 22, *every new partner required significant code development, especially to bring metadata into line.*)

How visible should the preservation archive be? What rights are partner organizations prepared to part with in order to accomplish preservation? (See p. 32, *HathiTrust is founded on public access and will not restrict access unless content bears legal restrictions.*)

What kind of governance structure do we need to set up a long-term success? (See p. 34, *if we develop a digitization preservation solution, we need tighter governance in terms of some authorized group to make decisions addressing some of the questions raised earlier.*)

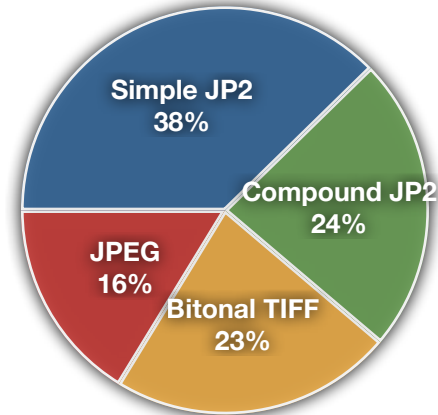
What we sent to HathiTrust

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 on the right.

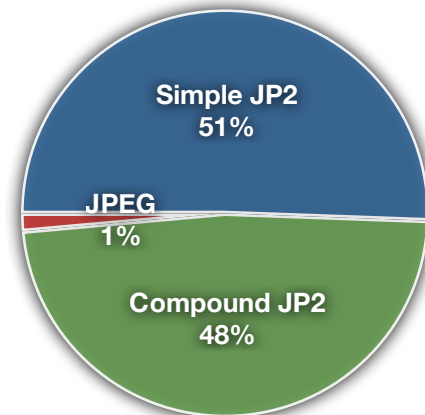


	Items	GB
Simple JP2	22,186	429
Compound JP2	13,844	407
Compound Bitonal TIFF	13,272	1
JPEG	9,575	12
Total	49,302	849

Number of Items Transferred



Amount of Data Transferred



Next steps

We are running through 4,000 image trials with both MetaArchive's LOCKSS network and OCLC's Digital Archive to help us find alternative answers to our questions. We will also evaluate the California Digital Library's microservices, Chronopolis, and DAITSS on paper and through phone calls. We are developing a matrix that addresses our questions and expect to be able to make a considered decision of the best way to proceed with digital preservation this summer.