

Minnesota Digital Library and HathiTrust



Prototype an Image Preservation Archive

5 April 2011
CNI Spring Task Force Meeting

John Butler, *University of Minnesota*

John Weise, *University of Michigan*

Eric Celeste, *Consultant, MDL*

Minnesota Digital Library, *Butler*

HathiTrust's Interest and Role, *Weise*

Prototype and Lessons Learned, *Celeste*

Minnesota Digital Library Coalition

- Conceived in 2001
- Early years – LSTA paycheck to paycheck
- Minitex (UMN) now the administrative home
- Signature project — *Minnesota Reflections*
- 62K images, maps, documents
- 120 cultural heritage institutions



MDL Access

“...not possible without you”

The screenshot shows the homepage of the Minnesota Digital Library's 'reflections' section. The header features the Minnesota Digital Library logo and the text 'MINNESOTA DIGITAL LIBRARY | minnesota reflections'. Navigation links include 'home', 'contributors', 'my favorites', 'preferences', and 'help'. A 'WELCOME!' message is accompanied by a scenic image of a lake and trees, with text stating that the site offers nearly 62,000 images and documents from over 120 cultural heritage organizations. A search bar is present with the text 'time' and a 'search' button. Below the search bar are sections for 'BROWSE' (by collection, topic, region) and 'TAG' (blackandwhite, blueearthcounty, etc.). A 'New Participants' list includes Bethel University, College of St. Scholastica, and others. A 'New Items' list includes an Oliver Mining Company Map Book and architectural drawings of Duluth's aerial lift bridge. A 'Coming Soon' section lists photos of Carleton College student life and letters from a WWI veteran.

home • contributors • my favorites • preferences • help

MINNESOTA DIGITAL LIBRARY | minnesota reflections

WELCOME! Minnesota Reflections now brings you nearly 62,000 images and documents shared by more than 120 cultural heritage organizations across the state. This site offers a variety of resources on Minnesota's history for researchers, educators, students and the public. We add throughout the year so check back often.

■ **SEARCH**
ADVANCED SEARCH

■ **BROWSE** ...by collection
...by topic
...by region

■ **TAG** blackandwhite blueearthcounty clothing
macalester macalestercollege man mankato men
minnesota mn oldmain outside postcard saintpaul
sepia students trees vintageclothing women 1910

■ **New Participants**
Bethel University, College of St. Scholastica, Cook County Historical Society, Mesabi Range Community College Library, Nobles County Historical Society, St. Croix Watershed Research Station, St. Paul Public Housing Agency, St. Scholastic Monastery, Synod of Lakes and Prairies (Presbyterian Church)

■ **New Items**
Oliver Mining Company Map Book, ca. 1926; Washington County surveyor's record book, 1847-1857; Architectural drawings of Duluth's aerial lift bridge, ca. 1902-1930; Souvenir booklet from Virginia, Minnesota, ca. 1925; Dakota language Christian missionary materials, ca. 1870s - 1920s; Oral histories from the Jewish Historical Society of the Upper Midwest; Minnesota Official State Highway Maps, 1950 - 2009

■ **Coming Soon**
Photos of Carleton College student life and activities, ca. 1895-1899; Letters from Trondhjem-area WWI veteran Ole Berg, ca. 1910 - 1921

MDL Access

“...we’ve got ours and we’re keeping it”



MDL Preservation

...something we can all agree upon

- Viewed as *common infrastructure*
- Economies/imperatives of scale
- Access need not be affected
 - but might it be?
- Attracting broader interest
 - including public radio & television
- A gift from Minnesotans
 - Arts and Cultural Heritage Funding
- A digital MLAC for the state

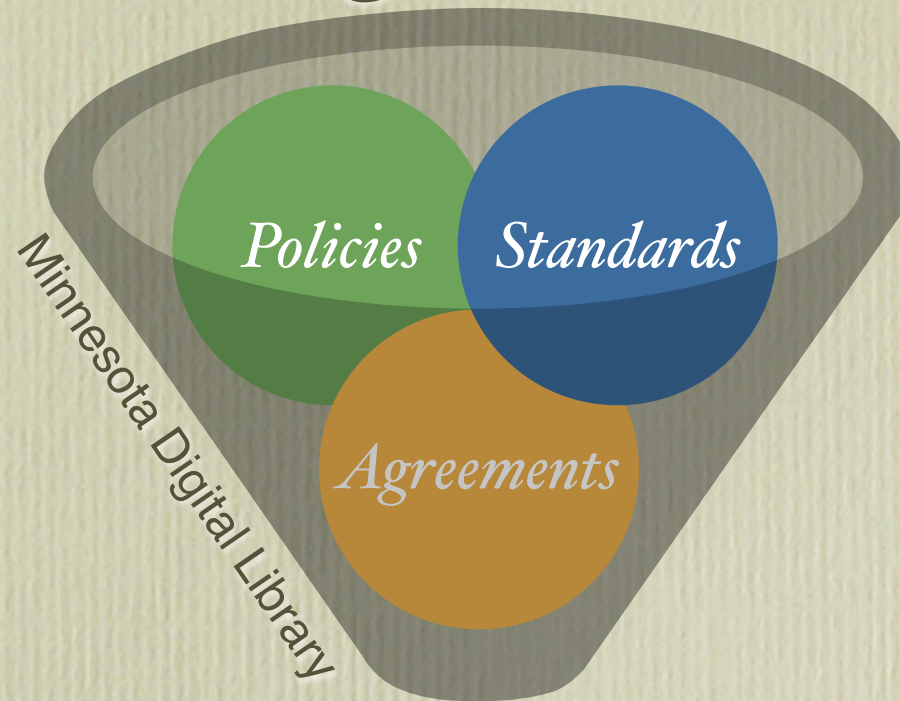
MDL Aspirations

State-wide Digital Preservation Services

The Minnesota Digital Library (MDL) seeks to explore a **common infrastructure strategy** that will bring the state a **significantly enhanced capacity** for preserving and accessing its cultural heritage.

To move the discussion from the hypothetical to the practical, we should begin **building a prototype**. It should be **collaborative**, meeting the needs of the primary partners (MDL, UMN, MHS, Minitex) and **extensible to other partners** (e.g., MPR, TPT, county and local historical societies).

Nesting Consortia



HathiTrust

Mission

- Access
- Preservation
- Research foci

Policies

- Governance
- Rights
- Costs

Formats

- Books
- Journals
- Experiment:
images & audio

Standards

- File Formats
- Procedures

HathiTrust's Interest and Role



Mission of HathiTrust

Contribute to the common good by...

- collecting,
- organizing,
- preserving,
- communicating,
- and sharing

...the record of human knowledge.



Preservation Philosophy

- Maximize...
 - **partner contributions**
 - **use of available resources**



Long Term Functional Objectives

- TRAC compliance
 - Robust discovery mechanisms
 - Open service definition (APIs)
 - **Support for formats beyond books & journals**
 - Data mining tools
-
- All functional objectives... <http://www.hathitrust.org/objectives>



HathiTrust's Interest

- Help MDL find a solution
- Explore image support
- Leverage MDL resources
- Empower MDL in the process
- Draft ingest specifications
- Establish a viable model for support of MDL and similar preservation cooperatives.



Design

- Operational logistics at the scale of HathiTrust have led to solutions that favor:
 - consistency and standardization
 - simplicity over complexity
 - practicality over ideology



HathiTrust's Role

- Scoping and guidance
- Lower barriers
- Raise bars
- Learn, grow and eventually provide
 - Object ingest specifications for images
 - Object preparation tools, esp. validation



All Together Now

- HathiTrust really is a collaborative effort.
- MDL and Michigan worked together under HathiTrust governance.

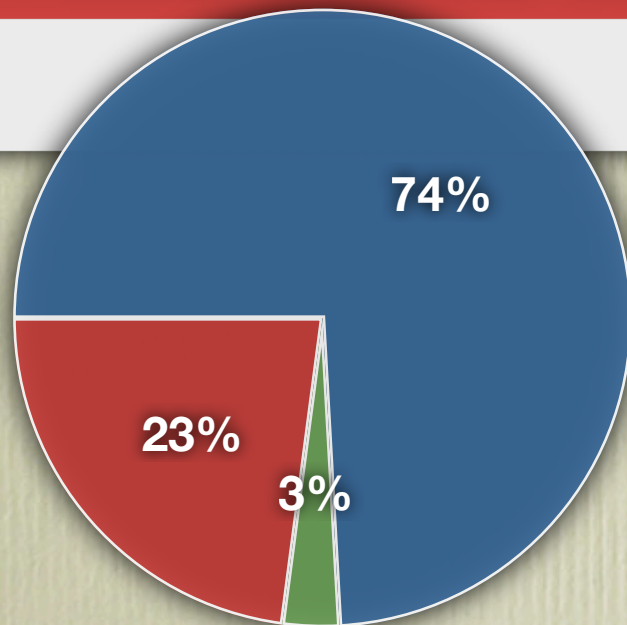


What did MDL actually
send to HathiTrust?



What MDL sent to HathiTrust

	Packages
Reflections Simple Contone	22,186
Reflections Compound Objects	888
Minnesota Historical Society	6,860
Total	29,934

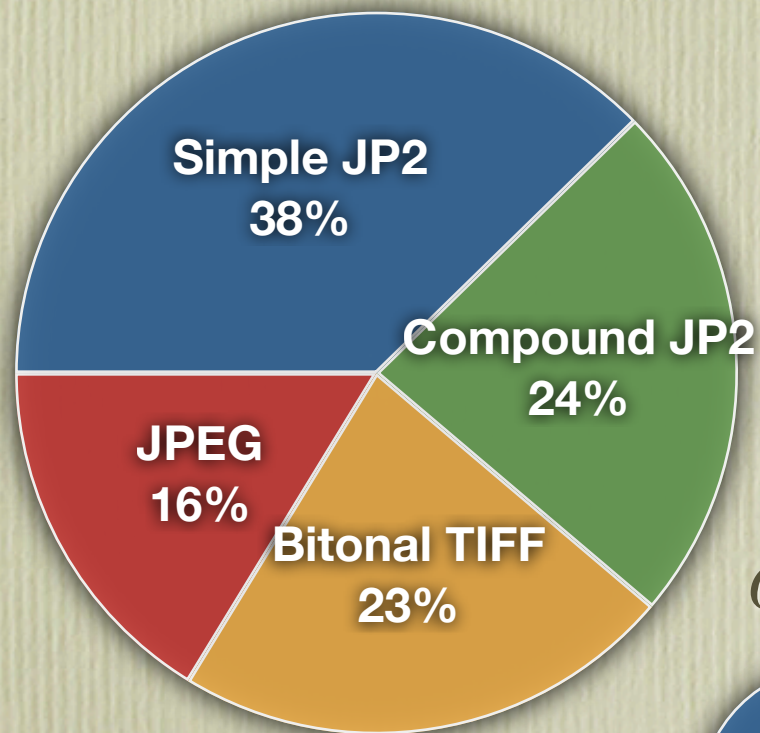


What MDL sent to HathiTrust

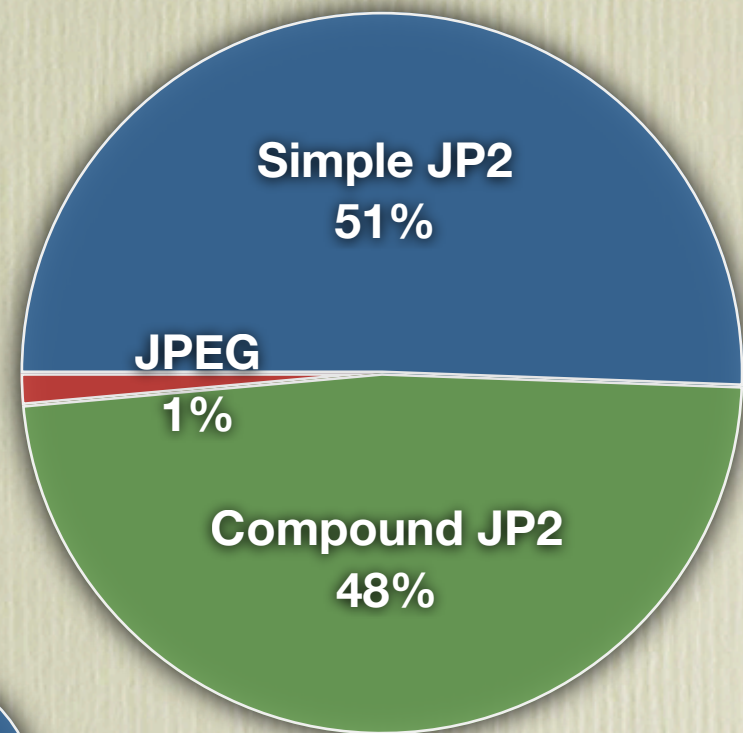
	Items	GB
Simple JP2	22,186	429
Compound JP2	13,844	407
Compound Bitonal TIFF	13,272	1
JPEG	9,575	12
Total	49,302	849

What MDL sent to HathiTrust

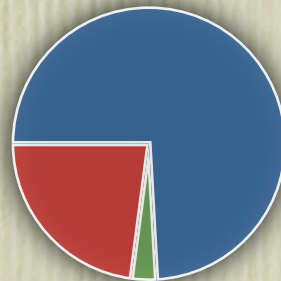
Number of items transferred



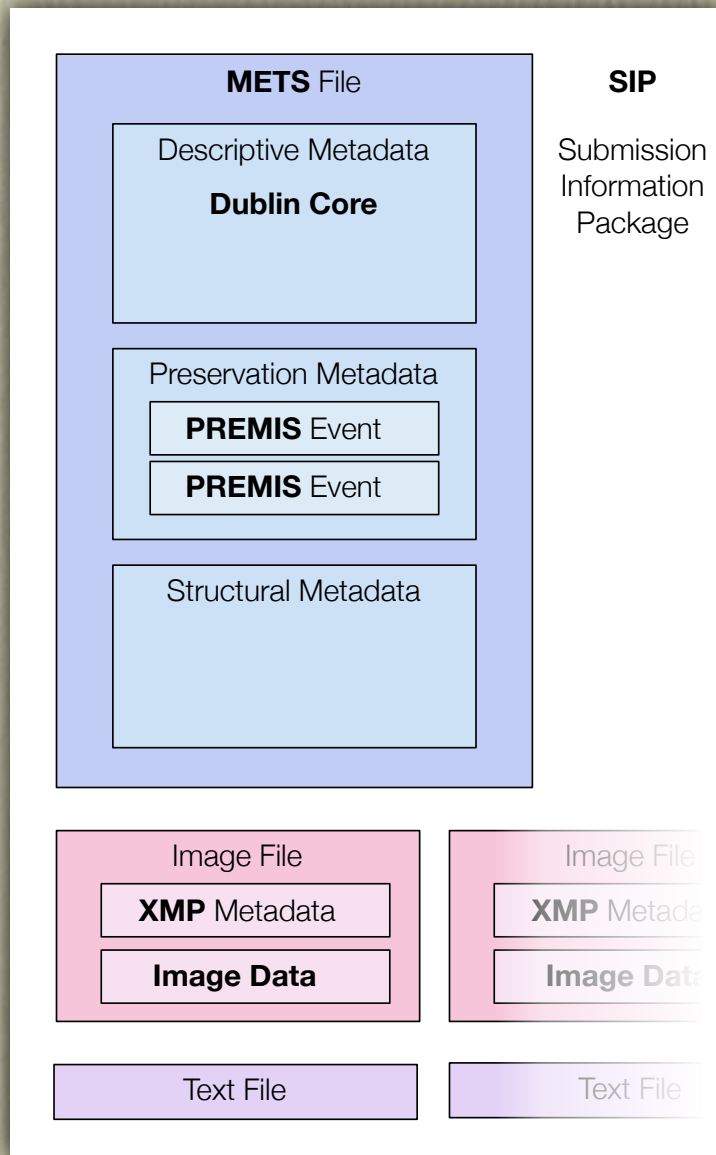
Amount of data transferred



Objects



What MDL sent to HathiTrust



An “object” sent to HathiTrust is a “Submission Information Package” and consists of many parts: a METS file with a variety of metadata, a set of image files, and a set of corresponding text files. The image files are the “items” sent.

What did we learn
from working together?





Get the full report at...


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



What is a master?



Where is the identifier?



SELF
PORTRAIT.

Metadata madness!

Trust us.





Who's looking?



No free lunch.

No free lunch.

	Time as Producer	Cost as Producer	Time as Aggregator	Cost as Aggregator
Programmer	4 weeks	\$8,000	14 weeks	\$28,000
Metadata Assistant	100 hours	\$2,000		
Manager	40 hours	\$2,800	30 hours	\$2,100
Totals		\$12,800		\$30,100

Next Steps for MDL?

- Trials with MetaArchive and OCLC Digital Archive.
- Paper and phone evaluations of a few others.
- Decisions in early summer about how to proceed.

Questions?

