

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

The screenshot shows the Minnesota Digital Library website. At the top, there's a navigation bar with links for Home, About, Projects, Digitizing & Metadata, For Educators, Tools & Training, and Contact. The main header features a green Minnesota state icon and the text "MINNESOTA DIGITAL LIBRARY". Below the header, there's a banner image of a scenic Minnesota landscape. On the left, a sidebar for "minnesota | reflections" includes a search bar for images, maps, and documents, a "Call for Proposals" section with a link to "Issue 7 Call for Proposals and Application Form", and a "MDL Listserv" section with a link to "Subscribe to the MDL Listserv". The main content area has a heading "Minnesota Digital Library News & Events" followed by several news items:

- Save the Date – Minnesota Digital Library 9th Annual Meeting** (March 8, 2011)
Minnesota Digital Library 9th Annual Meeting
Monday, June 6, 2011
8:30 AM - 3:30 PM
Continuing Education and Conference Center
St. Paul Campus, University of Minnesota
Watch for more information on speakers and registration.
There will be no registration fee. Participants will need to pay for parking.
- MDL Digital Preservation Project** (February 14, 2011)
During Fall 2010, the Minnesota Digital Library (MDL), using funds from the Arts and Cultural Heritage Fund, developed and implemented a digital preservation pilot project, to begin exploring ways to ensure that the content from Minnesota Reflections and other institutions, including the University of Minnesota and the Minnesota Historical Society (MHS), will always be accessible to Minnesotans.
From September through December, MDL staff and consultants worked with HathiTrust, a national shared digital repository, to develop a prototype project to add content to HathiTrust as a preservation archive. During the project period, nearly 10,000 images from Minnesota Reflections and 8,000 MHS images were shipped to HathiTrust.
→ [Read full article: "MDL Digital Preservation Project"](#)
- Minnesota Reflections Metadata Guidelines** (December 3, 2010)
The Minnesota digital library's newly revised Minnesota Reflections Metadata Guidelines are now available.
The new version provides better explanations of the metadata fields the Minnesota Digital Library asks contributing organizations to provide in their projects. The revised guidelines also contain advice on writing titles and descriptions and on assigning terms in many more categories than did our first version of the guidelines. They contain more examples and best practices. They also include extensive appendices to help explain and illustrate the metadata categories.
→ [Read full article: "Minnesota Reflections Metadata Guidelines"](#)
- MDL Introduces "steve," a New Online Tagging Tool** (December 3, 2010)
The Minnesota Digital Library has added a tool to allow tagging to many objects in its Minnesota Reflections online collection.


MDL Access

“...not possible without you”

The screenshot shows the Minnesota Digital Library (MDL) website, specifically the 'minnesota | reflections' section. The header includes the MDL logo and navigation links for home, contributors, my favorites, preferences, and help. The main content area features a 'WELCOME!' banner with a photograph of a lake and mountains. Below the banner are search and browse functions: a search bar with a 'time' dropdown and a 'search' button, and browse options for collection, topic, and region. A 'TAG' section lists various keywords. To the right, a sidebar highlights '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)) and 'New Items' (Oliver Mining Company Map book, ca. 1928; 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). The background of the main content area features a photograph of a misty lake.

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 time 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 vintagedressing 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. 1928; 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. 1918 -- 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

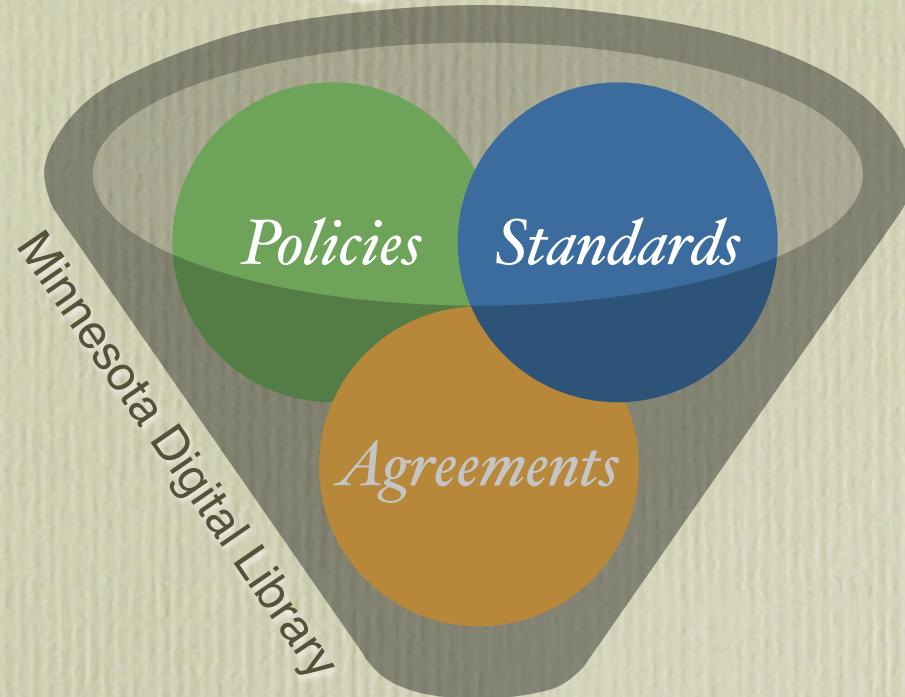
MDL Aspirations

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

<i>Mission</i>	<i>Policies</i>	<i>Formats</i>	<i>Standards</i>
<ul style="list-style-type: none">• Access• Preservation• Research foci	<ul style="list-style-type: none">• Governance• Rights• Costs	<ul style="list-style-type: none">• Books• Journals• Experiment: images & audio	<ul style="list-style-type: none">• 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.

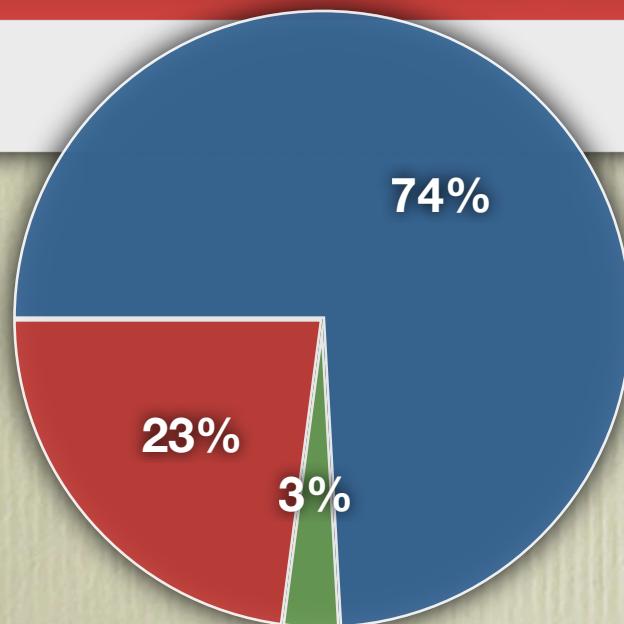


What did MDL actually send to HathiTrust?



What MDL sent to HathiTrust

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

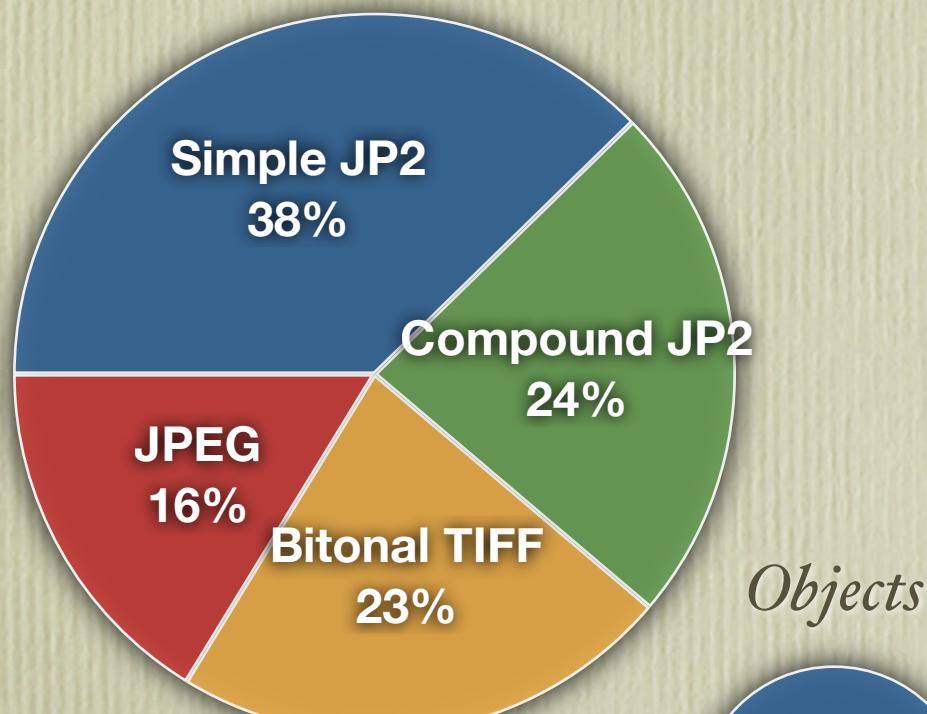


What MDL sent to HathiTrust

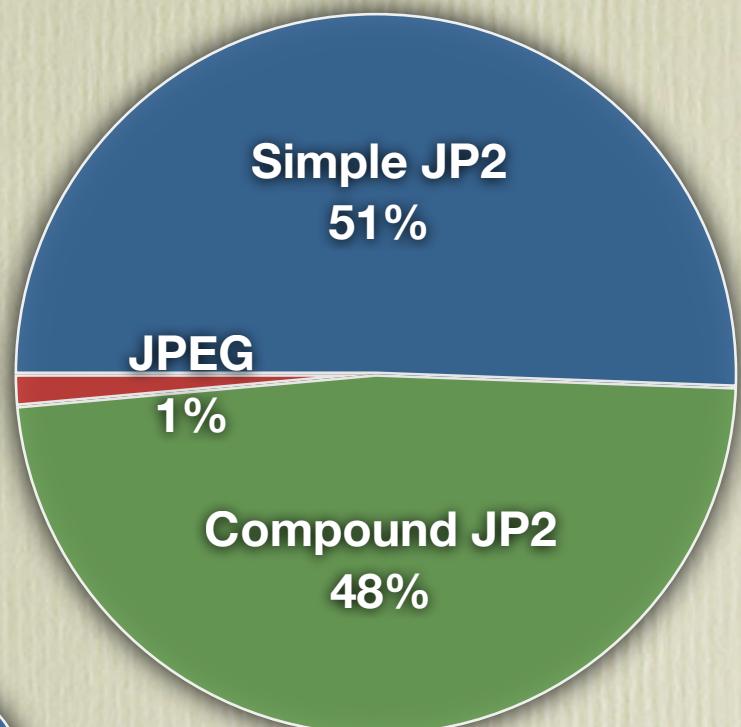
	Items	GB
Simple JP2	22,186	429
Compound JP2	13,844	407
Compound Bitonal TIFF	13,272	1
JPEG	9,575	12
Total	49,302	849

What MDL sent to HathiTrust

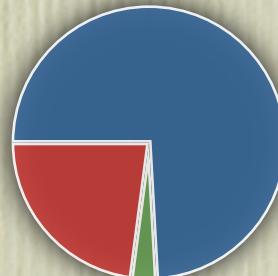
Number of items transferred



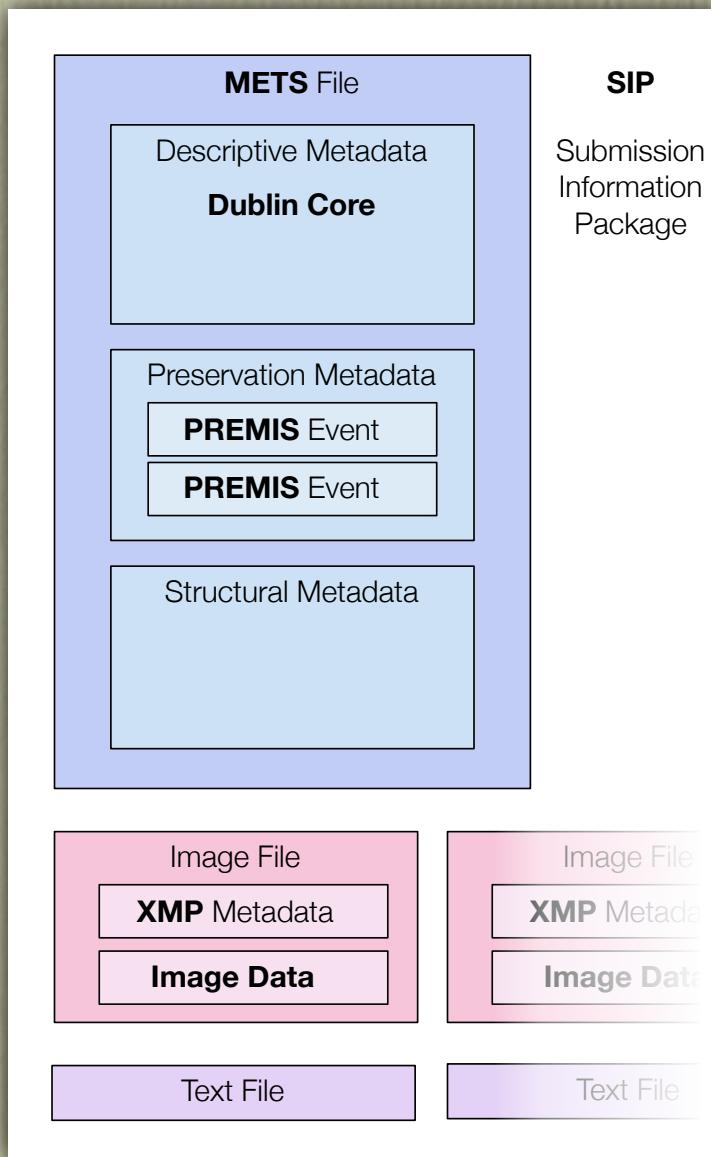
Amount of data transferred



Objects



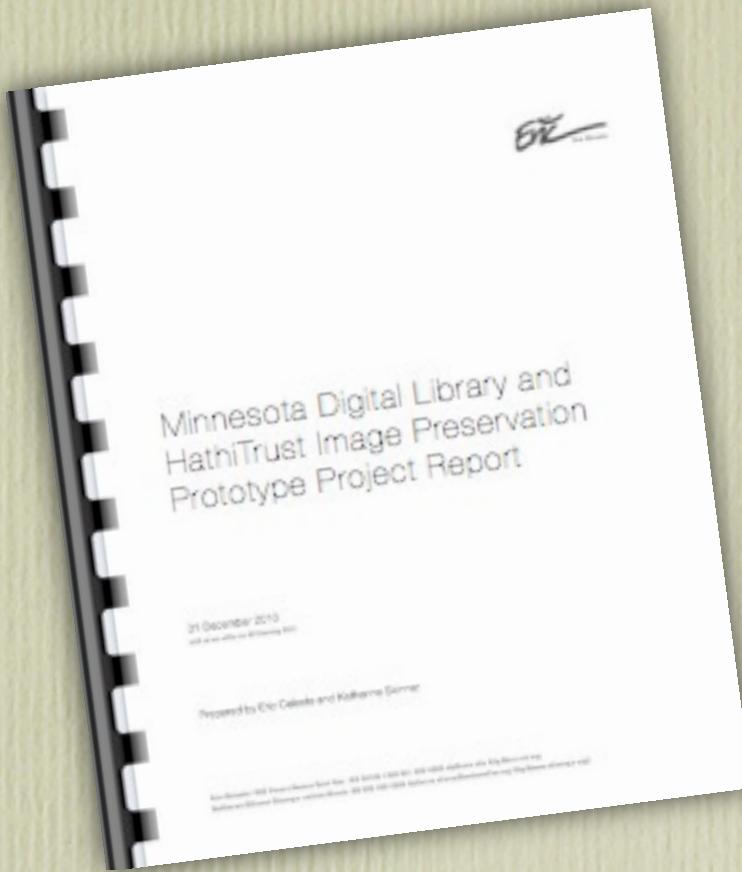
What MDL sent to HathiTrust



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?

A photograph of a man standing in a field of tall green grass under a clear blue sky. He is wearing a brown polo shirt and blue jeans, and is pointing his right index finger upwards towards the sky. The image is used as a background for a presentation slide.

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?

