

Minnesota Digital Library Coalition

Steering Committee Meeting

July 18, 2003

MINUTES

Attending: Mike Kathman, Bill DeJohn, John Butler, Michael Fox, Kristi Tornquist, Keith Ewing, Eric Celeste, Jason Roy, Marian Rengel

Location: MINITEX offices, Andersen Library, U of M

Budget Report (Marian): We still have more than \$23,000 to spend under the current LSTA grant. Spending or encumbering deadline is August 31.

Metadata workshop report (Eric and Jason): To be held from 10 a.m. to 3 p.m., Aug. 21, at the Olmsted County Historical Center in Rochester. Eric and Jason will present on metadata, technical issues of scanning, clarifying what digitization is. The workshop is targeting places that have started digitizing and working with metadata. Size: maximum of 30 attendees. MINITEX can handle registration.

Technical discussion: Plans for the next LSTA grant.

Eric, representing a sub-group that planned technical issues, presented a proposal (see attached) from the MDLC Technology Working Groups (Keith E., Jason R., John B., Eric C. and Bill Tantzen.

RECOMMENDATIONS:

- MDL would centralize storage of project metadata. The metadata would be sharable and held in one location. MDL will be a repository of thumbnail images and reference images.
- Creation of regional scan shops
- Metadata developed by image provider and scanners
- Architecture: MDL needs its own Tomcat/Apache server to host Web content
- Piggyback on U of M IMAGES database and Oracle services

BALLPARK COSTS:

Tomcat/Apache \$20,000

Oracle licenses: \$20,00/year

Refinement of IMAGES: \$50,000/year (personnel)

Customizing IMAGES skins: \$10,000/year (personnel)

Scanning workstation/server: \$5,000

Scanning costs: \$35,000/year (personnel)

Total First year: \$145,000

Ongoing annual costs: \$75,000

ASSUMPTIONS:

- 10,000 images added to repository each year
- Scanning would be done in a centralized location
- Does not include costs of metadata generation

- Does not provide for storage of archival master (recommendation from discussion was to have owner of material store the archival image)
- We do not treat signature project as a long-term asset to be forever preserved

DISCUSSION:

- Ballpark figures do not include hardware replacement or long-term recurring costs.
- We assume people who have the objects will generate metadata.
- Costs are loaded to the front.
- Michael F. recommended that we beef up output to more than 100-200 images a day.
- We could designate scanner at Stearns History Center as a regional MDL scanner.

Summary of June 2 Conference Evaluations (Keith): received a 4.3 overall (on a scale of 1-5)
Comments were all over the board

Signature project: We need a title

Some criteria for digitization projects:

- 1) About Minnesota
- 2) No copyright issues
- 3) Continuous tone images
- 4) Objects with “broad appeal” to others
- 5) Provide metadata that complies with MDLC guidelines prior to digitization

What image providers get out of this:

- 1) Free scanning – you get the archival images. Metadata is providers’ cost match.
- 2) Project director’s job: Consulting. Provide an appropriate metadata scheme

Minerva: This group is planning its third annual meeting Nov. 14 at the Weisman Art Museum at the U of M. John B., on behalf of Minerva, asked if the MDLC could contribute some of its remaining grant funds to supporting Minerva’s meeting. Specifically, he asked for between \$1,000 and \$1,500. The group directed Marian to consult with Jym Wroblewski on this possibility.

Discussion of spending:

How do work flow issues influence costs? Can we encumber grant money? Do funds have to be invoiced? Delivered? Paid? Marian will check with Jym on these issues too. Domain names? How much will parking several domain names cost? Can we do this?

Next meeting: 10 a.m. to 1 p.m. Friday, Sept. 5 at the MINITEX office.

Submitted by Marian Rengel

Attachments: Word table representing poster-size Post-Its from meeting
Technology Working Group’s proposal

POST-IT #1

	1 time	Annual
Project Director (full time)		\$60,000
Server (Tomcat/Apache)	\$20,000 (\$10,000)	
Domain name (currently \$100/5 years)		
Oracle licenses	\$10,000 ←	\$20,000
Server support		\$10,000
IMAGES refinement (1/2 [time] for 1 year)	\$50,000	\$10,000
MDL Skin	\$10,000	
Bridge costs (UofM IMAGES-SHM-MDL images)	?	?
Scanning Workstation	\$5-\$10,000	
Dedicated server	\$5,000	
Staff costs for scanning		\$35,000
[Total]	\$90,000	\$135,000
Minerva	[\$1,000-\$1,500]	

POST-IT #2

To spend:

\$23,000 by 8/29/03

\$4,000 by 9/30/03

\$125,000 [Round I]

\$175,000 [Round II]

mdl.net

mdlc.net

Below are the notes from two meetings of an MDLC technology working group. We will discuss these findings during tomorrow's MDLC Steering Committee meeting. See you then! ...Eric

MDLC TECHNOLOGY WORKING GROUP

On Monday 6/23/03 and Monday 7/7/03 Keith Ewing, Jason Roy, John Butler, Bill Tantzen, and Eric Celeste (in different combinations) discussed the technology requirements of the MDLC signature project.

STEERING COMMITTEE DECISIONS

We began by noting three decisions made at the last MDLC Steering Committee meeting: (1) adoption of the Western States capture standards, (2) adoption of the Western States metadata standards, and (3) agreement that IMAGES would be the foundation of the signature project system.

IMAGES is the digital repository developed at the University of Minnesota (Twin Cities) to hold digital image collections of the University. Much of its development and refinement has been funded by an IMLS grant to digitize and present posters from the World War I & II era. We agreed that as long as the signature project was aimed at collecting digital images, the IMAGES system was a good choice for getting us on our feet quickly. We spent the meeting trying to understand what the issues would be around adopting IMAGES.

METADATA REPOSITORY

We quickly agreed that the MDLC repository would centralize the storage of project metadata. This metadata would be sharable (via OAI), but would be held in one location for signature project purposes.

REPOSITORY OF THUMBNAIL AND REFERENCE IMAGES

We discussed what type of digital images would be required in the repository. This was a bit more complex than the issue of metadata, since the "master images" (the high-resolution TIFF images) could require massive storage to centralize. In the end we decided that "thumbnail" and "reference" images would be derived from the "master" images and would be required in the MDLC repository, but the master images themselves would not be stored centrally.

This decision was informed by the understanding that the signature project would really emphasize access to images rather than preservation of the

images (either preservation by offloading some wear & tear to the digital form, or preservation of the digital object itself). However, we did note that the lack of priority on preserving the master image might hinder some grant prospecting.

PRODUCTION PROCESS

We discussed process of producing digitized images and corresponding metadata for the signature project. It was felt that encouraging each participant site to do it's own scanning was probably not wise. There are just too many fussy details to the scanning process, and the temptation to just wing it with inferior equipment would be too great. Instead, we imagine there will likely be a few "regional scan shops" that can do the scanning for the participants. These shops would always provide copies of master/thumb/reference images back to the participant, but would also mount the thumb/reference images on the system.

The question of who would create metadata got a bit trickier. The consensus was that it made most sense to require the participating site to produce metadata for a given batch of images before those images were sent to a regional shop for scanning. The scan shop would then add the image metadata to the record begun by the participating shop. This is a very attractive model and appears to eliminate fears that a local site might not follow through on its part of the bargain. It would require some further development of IMAGES, such as the addition of a record "suppression" feature.

STEARNS HISTORY PROJECT

Keith described a c. 500 image project at St. Cloud that might serve as a pilot collection for the signature project. The scanning is well underway, though much of the metadata has yet to be transcribed from paper. St. Cloud needs to have these images in a usable database by the end of the summer, when the grant for the Stearns History Project runs out. We agreed to consider this as a driver for getting our MDLC repository off the ground.

GENERAL ARCHITECTURE

Bill and Eric reported that University Libraries staff took a closer look at the IMAGES system we've been planning to use of the MDLC effort and determined:

- (1) The MDLC signature project should provide it's own Tomcat/Apache server. This server would host the web

content of the project, including stylesheets, reference and thumbnail images, and any other pages or content associated with the project.

- (2) The project should piggyback on the Oracle services already provided by the Libraries for our University instance of IMAGES. This may require additional licenses of Oracle, but it is much less expensive than setting up a separate Oracle server. IMAGES is not ready to migrate to a database platform other than oracle for now.

These servers are likely to be sited with University Libraries servers in the Wilson Library with hardware/systems support provided by the Information Technology Services team of the Libraries.

If the cost of this architecture is prohibitive, we could consider offloading the "plain old web services" portion of IMAGES to a server housed at the University's Office of Information Technology (OIT). This would relieve the Tomcat server of the need to serve up images from a repository and might allow us to use less expensive equipment for that task. This option has not yet been discussed with OIT and we don't really know the cost impact of the idea yet.

BALLPARK COSTS

We discussed the various elements of the signature project and what they would actually cost us. These are very rough ballpark estimates, but they do give us a rough idea of the magnitude of these costs. Note that they are geared toward "sticker shock" rather than "wallet shock". In other words, we may not have to spend all this money, but we should at this stage anticipate that our costs could be this high.

- Hosting
 - of metadata, reference, and thumbnails
 - Tomcat/Apache server (V240 class machine w/360GB disk)
\$20,000
 - Additional Oracle Licenses
20,000
 - Server support staff costs
10,000
- Development

further Record Editor and ingest process refinement

Further refinement of IMAGES for MDLC (year one)

50,000

(further years)

10,000

- Customization/Skinning
designing and MDL skin and documenting process for futher
skinning

Initial design work for MDLC "skin" for IMAGES

10,000

- Production
scanning and cataloging of images

Scanning workstation

5,000

dedicated server for central location

5,000

Staff costs for scanning c. 40 images/day

35,000

First year ballpark cost of signature project via IMAGES

\$145,000

Ongoing annual costs of signature project

75,000

These costs assume a few things we should discuss further at the MDLC
Steering Committee meeting:

For one thing, we assume that the signature project will **target roughly
10,000 images** added to the repository per year of the project. This is a
relatively small number.

We further assume that the **scanning is being done in a central
location**. Setting up more regional scanning centers could be done and

would increase the target number of images the project could shoot for, but it would add further "Production" expense.

We **did not ballpark the cost of metadata generation**. In other words, participants would provide metadata at their own expense, then ship the items to a central location for scanning.

We also **do not provide for the storage or retention of the "archival master"** images, which are returned along with the objects to the owners.

For another thing, **we do not treat the signature project as a long-term asset to be forever preserved**. In fact, it has a rather explicit horizon of about three to four years. After that, some other costs come into the picture, such as the maintenance of the hardware purchased to run the first generation system. If the MDLC intends for the signature project to last for ages, then a more stable funding base than grants will have to be found.

#

*Eric Celeste / 612-624-4126 / efc@umn.edu / <http://umn.edu/home/efc>
Associate University Librarian for Information Technology
University of Minnesota (Twin Cities) Libraries*