

Digital Initiatives Steering Committee
Meeting of Wednesday, Oct. 7, 2009
8:30 - 10:30 a.m.
Thompson Library 150B

Attended by: Sally Rogers, Florian Diekmann, Mary Scott, Dan Noonan, Nena Couch, Amy McCrory,
Tschera Connell, Beth Black
Notes by Beth Black

Proposal for Google to scan selected Ohio Agricultural Experiment Station (OAES/OARDC) bulletin series (Sally)

Can go ahead with this, sent Google instruction book to Connie, still need to deal with barcoding these items, Connie working on options.

(1866 analytic bib records)

Way to move forward but work required on Connie's end next + mechanics of sign off

Google Book Scanning Project webcast on 9/2/09 sponsored by EDUCAUSE (Amy author of this portion of minutes)

**Amy's role here is to be the note-taker reporter as the one who could attend the session. She is not an expert.

This took place shortly before the Justice Department, on September 20, rejected the Google Books settlement with publishers and authors, so a lot of this no longer applies.
From the statement:

It noted that the proposed settlement "raises significant legal concerns" about whether it is "consistent with copyright law and antitrust law."

The department also raised the question of whether the societal stakes were too high to be decided by a private court action. "A global disposition of the rights to millions of copyrighted works is typically the kind of policy change implemented through legislation, not through a private judicial settlement," it said in the statement.

The Justice Department's recommendation does not mean the deal is dead. The department held out hope that the settlement could be altered enough to meet current legal standards, and it encouraged the parties to keep negotiating. It also went out of its way to talk about the greater good at stake.

Dan Clancy—Google Books
Jonathan Brand—Library Copyright Alliance (ALA, ARL, ACRL)

Clancy:

For works still legally in copyright by virtue of their publication date (post-1923). About 20 million books. Rights holders are represented by a group called the Book Rights Registry (BRR). The BRR does the work of locating owners of titles.

Rights holders may:

--Opt out of the settlement

--Remove books from scanning

--Select access models. Access models include showing snippets with links to WorldCat, Amazon, etc.

--Elect to charge, or not charge, for the works they own.

Brand:

Rights clearance is what has kept libraries from undertaking a project like this. While there are many benefits in digitizing these books and making them accessible, there are also problems:

--No competition. It will be an "essential facility" (Google has lead time advantage, others unlikely to enter market).

--No barrier to high prices if there's a monopoly.

--No equity of access, privacy, intellectual freedom.

--A few people are litigating the dispute; their decisions are binding on everyone.

The copyright problem of orphan works should be handled by legislation, not a class action suit and a private settlement.

Clancy:

Google is using an "algorithmic pricing model." It simulates how a real market for books would behave, finding "the right price." They think this will allow the market to set the price.

For libraries, there will be a "robust pricing model." For university libraries, pricing will be based on FTE model (number of students). Clancy says ILL expensive, so Google model is competitive.

Brand:

Suggests that FTE model is the same as that used for journal subscriptions, could result in similarly high prices.

LCA is not trying to stop the settlement, but to influence interpretation of it. They also want academic authors represented on the BRR.

Clancy:

Metadata problems (Geoffrey Nunberg, UC Berkeley School of Information, article): Clancy blames Union Catalogs, says Google's full text search is uncovering catalog problems. Nunberg actually demonstrated that this is not true; many problems arose from Google's use of bookstore categories, not library classifications.

Jonathan Brand's and Dan Clancy's PowerPoint slides are available here:

<http://www.educause.edu/Resources/TheGoogleBookScanningProjects/178822>

DRMC update (Nena)

Report from Aug and Sept meetings

Educational Technology Division is current favorite name of umbrella org for OL

- Sally noted they are building huge new infrastructure and hope it to be ready in next several months

- John frustrated by lack of resources so he is using Amazon cloud storage for 5 production instances of DRC
 - o Dan just named to task force to create guidelines for OSU about uses of cloud computing on campus
 - o Tschera asking about comment that we needed to keep OSU archival material on campus
 - o Sally will be attending Cloud Computing break out session at 10/22/09 CIO Advisory meeting
- Staffing issues at OL continue even when hardware ones are resolved, especially since ETD looks to DRC to provide these services
- OL interested in having Kyle's batch load software based on his demo
- Nena shared Anita's response to letter DRMC sent to her with some concerns

Some discussion of the preservation perspective of the DRC: do they restrict formats? Do they follow preservation standards? Do they have migration plans? Nena said that they will continue to ask questions.

Nena asked that we replace Kyle with another technical liaison on the DRC because his input was invaluable. Beth and Sally responded that we will once we have a permanent employee hired for that position.

OSU DRC instance

- OL not ready to give it to us
- We aren't sure what to call it
- Some discussion if we should move the 3 collections which are currently in the OL DRC to the OSU DRC.

Amy asked about command line and technical access to OL systems because the restrictions are getting stronger, specifically discontinuing ftp access for contributors. Sally has added it to list of items she will bring to an OL Tech Advisory Board meeting. This has implications for Brittle Books contributions to OL web books as well.

Improvements for the Mills Atlas on the OhioLINK pages (Amy)

Collection is in the DRC: <http://drc.ohiolink.edu/handle/2374.OX/>

Collection contains material by Ohio County

- Currently files appear in order they were entered into the system
- Can browse by title but still not quite right

Images offer zooming with high res images using Zoomify, making it easier to read online than on the originals

- Need way to make zooming window larger; perhaps pop a window?
- Keep the zoomify commands

Will these issues be rectified? Amy not sure if it will be fixed but she did tell John. Group brainstormed possible solutions. Amy asked for help and Tschera suggested that she work with Maureen Walsh.

Presentation approach used by the Internet Archive to "turn pages" for books in digital form (Amy)

Amy demonstrated

- Current OSUL delivery method: Link to ebook in catalog goes to PDF reader in the browser ; include OCR so searchable

- Complicated when books are really large (50 MB or 600 pages) and it gets split into sections (generally 4 pieces)
- Alternative option: Internet Archive's Flip book software
 - Read online with a page turner
 - Includes searching and zooming
 - Including these formats:
 - PDF
 - b/w PDF
 - djvu (jpg2000 based)
- Amy would like to offer this for our users. Led to discussion about who we would ask (OL?) and priority for putting this on our servers. Some options discussed:
 - Load page turning software to our servers and take over delivery and storage of ebooks here
 - Send our books to Internet Archive
 - Ask OL about future of web books area
- Should we continue to digitize with Google doing this?
 - (Amy) Yes because Google not doing fragile items. They check Google books first and don't scan things that are already done.

Migration Planning (Dan leading discussion)

This was a topic we identified as an important one for discussion when we planned for the year.

How to frame the decision making process for migration?

Should we set up a time each year that we look at the formats we have and those that are coming up? Or do we wait for the tools to change or other catalysts in the environment?

We need to be setting aside money for this. Sally noted that in order to get funding, we need to put it in terms of the risks that will be compelling to funders. Tschera notes that we need to talk also about our funding model in general for this work.

This planning must also include the cost of storage replacement + regular increases in storage capacity.

Types of migration for which we should plan:

- Refreshing hardware
- Migration of formats
- Migration of repository software

Next actions:

- Identify the number of items we have in each format (Tschera or Beth)
- Review of the supported formats list in KB (later – could it be taken up by the PDF/A working group?)
- Dan will flesh out an outline of a migration strategy to start

Updates:

Lantern

Paper issues already scanned: 1881-1882, 1885-1892, and 1959-1997.

Issues currently online: everything from that group up to 1982.

Microfilm issues, currently being scanned: 1883-1884, 1893-1924, 1925-1927, and 1928-1941.

Remaining paper issues, to be scanned in coming months: 1924-1925, 1927-1928, and 1941-1959.

(If some of the years seem repetitive, that's because the volumes are dated midyear-to-midyear, e.g., June 1928-June 1929).

The remaining work will cost \$17,350 more than the \$200,000 total funds allotted for the project.

Mapping and Charting Lab Tech Papers – no update

KB Toolkit Update (Tschera)

Will be used for a brown bag for the Subject Librarians being held soon and will be online at the SRI web page. Tschera is still working on a better place for this with the development of the KB Center website with Beth.

New Resources available:

- Captivate video describing items in toolkit
- packaged intro to KB PowerPoint
- pool of PPT slides for customized KB talks

Written agreements with KB communities (Tschera)

Started with the electronic publishing initiatives and still working on these agreements; they are proving harder to write as trying to accommodate the needs of different communities. Tschera will share more at the December meeting.

LITA Conference (Beth)

Text sent in email to group

IRODS (Dan)

Supplement to written report Dan provided to group via email.

IRODS is a tool to help with preservation and management of data

- could be used to track original and derivatives
- distribution across multiple repositories
- to help with migration and preservation planning
- using for website harvesting

Requires programmers with advanced language skills (C+, Java, Perl)

- University might have resources to implement but Libraries probably do not at this time