Digital Library Services TF : Report to CDC

Mass Digitization Collection Advisory Group (MDCAG) Status Report to CDC
September 21, 2007
Robin L. Chandler on behalf of MDCAG
Introduction: In accordance with its charge, MDCAG has developed "an internal process for the review, identification and selection of collections for scanning across the UC Libraries." The group has drafted goals, selection criteria (principles for selection), a collection proposal template and a set of procedures and workflow document for the identification, selection, review and approval of book collections for scanning. MDCAG has sought to balance tensions between the UC Libraries desire to digitize specific high value collections with the needs of the two mass digitization projects for large volumes of content. To address these sometime conflicting needs, MDCAG proposes a three-tiered model to identify materials supporting high volume scanning and identify collections important to UC to scan.
Three-Tiered Model:

- Google: Considering the high-volume demands of Google, (15,000 – 18,000 books per week), MDCAG believes it is best to identify campus libraries "ready-to-participate." MDCAG concluded that Google's needs for continuous high-volume scanning approach was not-well suited to a granular proposal process. Identification and selection of collections for the Google project is better suited to discussions with ULs and AULs in consultation with their local collection experts to advise as to which libraries are eligible to participate.

- Microsoft: MDCAG believes the proposal process described above is best suited to the Microsoft scanning project (2,000 books per week). MDCAG envisions a process where book collections can be proposed for scanning by UC librarians in consultation with their AULs for Collections. Proposals meeting the immediate criteria for the Microsoft project would go into the mass digitization queue approval process, where the Collection Development Committee (CDC) writ large and MDCAG (should its role as a group be formalized in the long-term) would be consulted and approve proposals.

- Digital Projects Registry: Proposals outside of the scope of mass digitization flows would be entered into the Digital Projects Registry (database). This process could serve as the mechanism to surface projects that have no mass dig funding, or do not meet criteria (lack cataloging, etc) but are worthy of considering for digitization by the UC Libraries. Collections determined to not meet the criteria for mass digitization, could be considered using the Tier 2 proposals model used for licensed content. Note that for a Tier 2 proposal there is a lead proposer that works with the AULs for Collections at other campuses as appropriate.

CDC Feedback Requested:

1. Comments on the Three-tiered model and thoughts about specific questions:
   - seek advise/ guidance from the CDC about their willingness to pursue digitization of their campus materials by the Google.
   - we can take advantage of the current two mass digitization sponsors Google and Microsoft by supplying materials that meet their stated criteria. If we solicit proposals beyond those criteria, are we raising false expectations across the campus libraries?
   - should we devote time and energy into the development of a proposal process to identify worthy non-book collections, which have other needs or require further work before they can be digitized. If we identify those, should we create a digital projects registry, where projects can be viewed systemwide, to identify synergies and needs? These could be categorized as Tier 2 proposals.
   - If we create a systemwide digitization proposal registry, who maintains it?

1. Comments on the ULs 9/13/07 discussions

Summary of Chandler's Report to University Librarians (ULs) on 9/13/07
On 9/13/07, Robin Chandler gave a report to the ULs about the status of the mass digitization projects. This included a brief overview of MDCAG's work to date. In summary the UL's learned MDCAG's work included:

- developing collection principles for mass digitization, and

- identifying a process for proposing campus collections for digitization in both the Google and Microsoft projects.

MDCAG's draft principles are applicable to all mass digitization projects, but MDCAG has focused on a process for Microsoft collections where there are fewer restrictions on sharing content for use, and in-house scanning facilities supports special collections. The advisory committee is reporting on its work to the CDC committee on September 21, 2007 seeking feedback on work to-date before finalizing their proposals and recommendations.
MDCAG also asked for feedback from the ULs regarding their interest in the group developing recommendations about a Digital Projects Registry (database) for supporting strategic planning about digital content projects systemwide. Should the ULs express interest in this approach, MDCAG will discuss the idea further with CDC to scope the planning effort.
<u>ULs response on MDCAG progress</u>

- The ULs approved the idea of identifying "ready-to-participate" libraries for the Google scanning projects.
    1. ULs took the action of approving discussions between CDL, Google and three campus libraries: UCLA, UCSC, UCSD to consider the Google project expanding to those libraries. This would be phase one; phase two would bring other UC Libraries into this discussion.

- The ULs grasped MDCAG's recommendations to focus the collection proposal process on books for the Microsoft workflow and thought it appropriate to have determinations about library participation in the Google project be informed by discussions with ULs and AULs.

- While expressing concerns about administrative overhead, the ULs were willing to have MDCAG continue thinking about the Digital Projects Registry concept with guidance from CDC and report back to them. They'd like a report to include:
    1. some thought about business requirements (who/what would be required to sustain such a process)
    2. scope of registry, i.e. would this include collections beyond the library including UC museums
    3. how would it intersect (or not) with OCLC's Digital Registry as we are now looking at a closer relationship with OCLC through WorldCAt etc.