Regional and National Cooperation on Legacy Print Collections
Boston Public Library, January 15, 2010
Meeting Notes

After welcoming remarks from meeting sponsor LYRASIS and participant introductions (see list of participants at the end of these notes), the group heard short reports as follows to augment the updates received in advance of the meeting (Kieft email of 1/12/10). Given that the reports from the meeting are long, I am summarizing results here at the beginning and providing a TOC of the reports.

1. The group generally endorsed the provisions of CRL's planning document and complimented CRL on its clarity and ambition. Several consortia expressed interest in working with CRL on the steps outlined in the document, and other interested groups should contact Lizanne (lpayne@crl.edu). It is likely that a meeting of interested parties will take place in Chicago in April in tandem with ICOLC.

2. Ellen Dunlap, AAS (EDunlap@mwa.org), reminded participants that she is looking for libraries to help test Clarence and Edna.

3. We will have a meeting at Annual, sponsored again by LYRASIS and again on Friday morning (June 25).

Outline of meeting notes:

A. Reports from consortia
    1. Western Regional Storage Trust
    2. Cloud Library
    3. LYRASIS
    4. American Antiquarian Society
    5. Ithaka S+R
    6. ASERL
    7. CIC and GWLA
    8. OhioLINK

B. Presentation and discussion of CRL planning document

C. List of participants

****

A. Short reports

1. Western Regional Storage Trust (WEST).

Bob Kieft and Lizanne Payne reported on the first meeting of this nine-month project funded by the Andrew W. Mellon Foundation. The following text is a meeting report written by Ivy Anderson.

For two days in early January, prospective partners met in Oakland, CA to discuss plans for the Western Regional Storage Trust (WEST), a distributed retrospective journal repository for the Western region of the United States. Participants represented university libraries in California, Washington, Oregon, and Arizona; special libraries including the Getty and the Huntington; and consortium leaders from the Orbis-Cascade Alliance, Greater Western Library Alliance (GWLA), and the Statewide California Electronic Library Consortium (SCELC). This meeting was part of a 9-month project planning effort funded by a grant from the Andrew W. Mellon Foundation.

The goal of the meeting was to explore selection criteria, risk management, and validation standards, and to select one or two preferred business models with operating plans for further development. Emily Stambaugh, Shared Print Manager at the California Digital Library, led discussions about selection criteria and risk management principles for print journals. Lizanne Payne, project consultant, presented information about other components such as archiving location (facility type), disclosure system(s), access/delivery policies and system(s), ownership/retention, financial support, and organization structure. She also presented several alternative business models which outlined various approaches to implementing a regional depository network.

Through extensive discussion and a couple of consensus-building exercises, the following preferences emerged:

1. Priority for selection:
   - Widely-held journals with electronic backfile and post-cancellation rights – these holdings promise the greatest efficiency gains in space utilization and collection management costs with the least impact on user access, and would also contribute to a national network of optimal copies.
   - Print-only journals with significant duplication within WEST (including titles widely-held at the network level as well as titles primarily concentrated within the region) – both for the efficiencies outlined above and to preserve archives of regional importance
   - JSTOR journals (i.e. to complete runs held in existing archives)
2. Preferred business models
   - Based at multiple storage facilities (or other secure library locations)
   - Based at multiple storage facilities (or other secure library locations) and also at distributed campus libraries

Over the next several months, the group will develop more detailed plans based on these outlines and on some collections analysis, for presentation and selection of a final business model at a followup meeting in June 2010.

Some of the key principles that were noted during the meeting are:
   - There is a big economic sweet spot, don't obsess over details around the edges

- Libraries are moving away from the hybrid print/digital model, into a largely-digital model (i.e. attempting to eliminate dual formats where possible and appropriate)
- Collective management of legacy print achieves real economies of scale
- Spend as little as possible on preserving legacy print that is in electronic form (because it has a limited life span) or that is widely held – an 'optimal copies' approach should be sufficient
- Consider national context to determine the optimal number of copies to retain regionally and nationally
- A coordinated approach at the regional and national level has strategic value in discussions with key stakeholders (i.e. faculty and administrators)
- Disclosure is critical for a cooperative effort to succeed

An informational website for WEST will be available on the CDL website <http://cdlib.org> at the end of January.

2. Cloud Library (ReCAP/NYU)

Michael Stoller was unable to attend the meeting but forwarded this "UNOFFICIAL STATUS REPORT: "TOWARD A CLOUD LIBRARY. A PROJECT TO EXPLORE THE POTENTIAL BENEFITS OF COMBINING LARGE-SCALE VIRTUAL AND PRINT REPOSITORIES AS SURROGATES FOR RESEARCH LIBRARY COLLECTIONS." Michael asks that we not spread this report around given that discussion is ongoing.

The proposal was funded by The Andrew W. Mellon Foundation as a planning grant for a three month research project to explore various models that combine the holdings of a digital repository and titles in an offsite print storage facility as an acceptable surrogate for onsite printed collections held by a research library. The goal was to research and identify the policy, procedures, logistics, and infrastructure necessary for the subsequent elimination of those printed titles duplicated in the university library, at a considerable savings of costs over time.

Rather than conducting an abstract study, New York University Libraries have been used as a paradigmatic client library for the project. The digital repository of the HathiTrust and the offsite collaborative storage and retrieval facility, ReCAP, are used similarly as operational, 'real-time' source examples for the research. Focusing on public domain material, the project strives to determine what service level agreements are required of ReCAP and HathiTrust in order to assume responsibility for serving as a surrogate for selected titles held by NYU libraries, and the associated cost modeling that ReCAP and HathiTrust need in order to make realistic, long term commitments as trusted, alternate sources of these materials. RLG/OCLC staff are conducting the research and developing a plan that addresses these issues, in order to make an initial determination of the percentage of NYU materials that fit the criteria for low use storage that are covered by ReCAP and the HathiTust depository, and to calculate what such a new arrangement might save (or not save) NYU balanced against the costs to ReCAP and HathiTrust.

Calculations by RLG/OCLC have sought to identify the overlap between the holdings of NYU Libraries, ReCAP and the Hathi Trust. To date the project has identified some 600,000 titles held by both NYU and Hathi, amounting to about 28% of the NYU Libraries holdings. It has been difficult to pin down the precise overlap between NYU's holdings and those at ReCAP, as the latter does not have a distinct catalog but is composed of materials owned by the three partner libraries, Columbia, the New York Public Library and Princeton. But data uncovered thus far, based exclusively on the Columbia ReCAP holdings, indicates some significant overlap between NYU and ReCAP, but only c. 7% of titles held at NYU are duplicated at both ReCAP and Hathi.

Approximately 94% of those titles duplicated at all three institutions are out of copyright and therefore available presently from Hathi online. But the low level of duplication among the three has presented serious issues, as we consider a model that would enable a client library like NYU to withdraw significant numbers of titles based on their availability in both physical and digital forms from partners like ReCAP and Hathi. We have considered whether there are idiosyncrasies about the NYU collections, particularly their relatively small retrospective holdings, that may indicate it is an inappropriate paradigm for a client library, though comparative data suggest NYU is actually more typical than not. The data has also spurred conversations about conditions under which a library like NYU would consider withdrawal of materials based exclusively on the availability of a digital surrogate from an institution such as Hathi. There is general agreement that current research climates this would not make such arrangements viable at present, but the conversations have been interesting.

Even as we have been considering the overlap data, conversations have been underway regarding the likely contractual and cost structures that both ReCAP and Hathi would require in order to make any relationship of this sort workable, and we have also considered whether partnership with additional print repositories would yield a higher overlap rate for a client library, and whether such partnerships would be a possible alternative to the three-way structure contemplated in the grant proposal.

Meetings are continuing between the partner institutions and RLG/OCLC, as we explore answers to the questions posed by the project and the data it has generated to date.

3. LYRASIS grant proposal to IMLS

Cathy Wilt reported on the grant proposal, which seeks funding from IMLS to support a national workshop in the fall of 2010 that would begin to address the question of monographs, given that other projects are well on their way to national and regional programs for legal material, journals, newspapers, and government publications. The following summary is the abstract from the grant proposal.

LYRASIS proposes a six-month planning project to explore development of a national framework for collaborative archiving and storage of print monograph collections. With assistance from its planning partners – the Center for Research Libraries (CRL), the California Digital Library (CDL), and the Committee on Institutional Cooperation (CIC)

– LYRASIS will define the characteristics of a national, collaborative model for retention of print monographs and identify issues that need testing or research to confirm the framework. The planning initiative will include organizations from throughout the U.S. in formulation of the collaborative framework. This collaborative framework has the potential to improve long-term preservation and storage practices for printed monographs on a national level, to establish a cooperative infrastructure that will result in efficiencies and savings for libraries, and to coordinate storage practices with digitization activities to ensure cost-effective archiving as well as reliable long-term access. As collaborators work together to form the national framework for print monograph storage, partners also plan to identify organizations able and willing to participate in research and/or test projects to refine the framework, for which a proposal would likely be submitted to IMLS following the planning grant project.

4. American Antiquarian Society

Ellen Dunlap rehearsed for the group the history of AAS, whose purpose is to preserve US publications printed pre-1876 and to make them available for study onsite and by microform reproduction and digitization. AAS is a print repository first and foremost, and in 2001 they opened a cold, high-density storage facility attached to their building with a capacity to double their holdings, which means room for everything printed in the US that they now do not have. AAS has specialized in creating high-quality metadata for US imprints with support from NEH. They have created a protocol called Clarence for issue-level information about newspapers and have finished testing it; the next step (which AAS plans to accomplish this summer) is to load data for AAS newspaper holdings, which number about 2 million issues. They have a similar tool for other serials called Edna, which EBSCO is going to help populate over the next three years as they complete the digitization of all pre-1876 periodicals in the AAS collection (more than 6,000 titles). AAS digitizes with many vendors and is working on POD. They receive many questions from libraries and historical societies about joint archiving and digitization and are happy to consider taking in any materials which are not already represented in the AAS collection.

5. Ithaka S+R

Roger Schonfeld updated the group on the reception of Ithaka's "What to Withdraw" paper and invited comment on it. ASERL has formally endorsed it. Most of those who have commented on the paper have been enthusiastic; libraries want to take action based on the kind of scientific analysis the paper offers. (A critique of the framework proposed by the paper has been made by one individual from the point of view of the "self-authenticating" nature of a print collection vs. the need to authenticate an electronic collection against print.) As part of this work, Ithaka has released a tool to help libraries make decision about retention of print serials in the national context for titles that have been digitized. Both the paper and decision tool are linked from http://bit.ly/BsIUE.

Ithaka S+R also released in December a report on the Federal Depository Libraries Program, with recommendations for its future as shaped by opportunities for expanded

user access associated with digital dissemination and the digitization of historical collections, while ensuring preservation and improving incentives associated with library participation (http://www.ithaka.org/ithaka-s-r/research/documents-for-a-digital-democracy).(http://www.ithaka.org/ithaka-s-r/research/documents-for-a-digital-democracy).

6. ASERL

Judy Russell distributed a handout and commented on the work of an ASERL "Deans Task Force on FDLP Issues." Her handout is copied in

Members: Judy Russell, University of Florida (chair); Bill Potter, University of Georgia; Julia Rholes, University of Mississippi; Bonnie MacEwan, Auburn University; Larry Boyer, East Carolina University; Lance Query, Tulane University; John Burger, ASERL (staff liaison)

Background

In November 2009, the Association of Southeastern Research Libraries (ASERL) convened this group o better manage the very large regional collections within ASERL while sustaining or improving services, including establishing a common and streamlined disposition process for federal documents.

Assumptions

Any decisions and actions will be in compliance with 44 U.S.C. and GPO policies;
The federal documents collections held within ASERL are a regional asset, not just assets for the holding library or for the state where the collection is located, and should be managed accordingly;
No library will be pressured to surrender Regional status; and
No library will be pressured to dispose of documents that are valuable to their institutions.

Goals

To plan collectively for management of federal documents held by ASERL libraries, in collaboration with all of the Federal Depository Libraries in our region;
To evaluate existing models for shared retention and disposition of documents collections (e.g., Indiana, Nebraska) to determine their viability across a larger number of FDLP libraries such as ASERL;
To gain consensus on a regional (ASERL) policy/best practices for disposition of federal documents that will simplify and streamline the process and address collection management regionally;
To incorporate the management of federal documents into the planning for collaborative print storage among ASERL members; and

To determine how the findings of Ithaka's report "Documents for a Digital Democracy: A Model for the Federal Depository Library Program in the 21st Century" can inform our regional planning.

7. CIC and GWLA

Mark Sandler reported that, like ASERL, CIC is working toward regional management of documents. As encouragement and support for a collective approach to managing print documents, they have entered into a partnership with Google to digitize them. The U of Minnesota has sent 86,000 duplicate documents for digitization by Google, to be accessible in BookSearch and the HathiTrust Digital Repository. Approximately 80,000 of the Minnesota volumes--and their corresponding bibliographic records-- will be made accessible via Hathi in the next several weeks. The remaining Minnesota volumes will then be reviewed for quality threshold issues, fold-outs, in-copyright inserts, etc. This review process is intended to gather information about the digitization workflow so as to optimize quality and efficiency as the project proceeds. Penn State University is currently sending a group of 25,000 items to Google for digitization, and other CIC universities are queuing up larger bodies of Federal documents for shipment in the summer of 2010. Over the next several months, many bibliographic issues need be addressed to allow our universities to get beyond already cataloged and barcoded content. A CIC working group has been tasked by the Library Directors to explore options for linking print content at one or more of our institutions with records that might exist elsewhere.

Public domain Hathi content is freely accessible to users worldwide. An API exists to link users to Hathi digital files through their local library catalogs, but at this time full MARC records have not yet been made available for ingest. As for the Digital Files themselves, Google has agreed that Hathi can make these accessible to GPO at such time as the agency might want to manage and distribute the content directly.

As a corollary to these documents projects, Joni Baker of GWLA reported on the TRAIL-Technical Report Archive and Image Library. This is a collaborative project to digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975. The Greater Western Library Alliance (GWLA - www.gwla.org) and the Center for Research Libraries (CRL - www.crl.edu) are collaborating on a pilot project to provide both digitization and long-term print archiving of these materials. Oklahoma State University is serving as the print repository, and the items are being cataloged (full MARC) at the University of Arizona, and digitized at the University of Michigan as part of their Google agreement. Ultimately, the items will be findable (is that a word outside of library land?) through Google, the University of Michigan catalog, and the Hathi Trust; the MARC records will be available in OCLC, and a separate searching interface will be developed by the University of North Texas.

Right now, it is just a GWLA project, but will be transitioning to CRL administrative control to become part of the Global Resources Network in the next 6 months or so--we are satisfied that we have the "proof of concept" figured out, and moving the project to CRL will allow a much greater number of interested institutions to contribute

collections.   CRL members should start receiving news about the project and how to join in March and April. I believe the TRAIL leadership group is also making provisions to allow non-CRL members to participate as well.

8. OhioLINK

Julia Gammon distributed a report about recent activities regarding OhioLINK print repositories. It is copied in here.

1. Statement of Agreement is in fourth draft and continues to be discussed.
2. Shared depositories catalog is in the planning stage. An OhioLINK implementation committee is working with III reps on profiling for a test database. The tentative (stress: very tentative) timeline is to begin loading records in late spring. With records to be loaded from 13 institutions, the catalog will be brought up incrementally.
3. Journal de-duping project – a small pilot has just concluded to test procedures and assumptions, and the results are being studied. Current plans call for retaining no more than two complete copies of a serial title across the five depositories. All things being equal (catalog load proceeds on schedule, budget situation doesn't worsen), the project should begin in earnest as soon as the Statement of Agreement has been approved.

B. Presentation of planning document by CRL

Bernie Reilly summarized the work CRL has done since our meeting in Chicago. CRL has:
- learned more about projects underway in GWLA, ASERL, CIC, etc. and talked with many institutions and consortia that are developing programs and facilities;
- looked at domain-specific programs (LLMC, US Agricultural Information Network), both of them longstanding partnerships to catalog and preserve digitally the literature of a field;
- developed a proposed with CDL to transfer its prospective archive of current journals to CRL, found a core of strong supporters but not broad consensus yet
- spoken with JSTOR and Serials Solutions about possible use of metadata
- evaluated large digital repositories to verify their ability to provide preservation in the long term in order to allow libraries to keep fewer physical copies (completed an audit of Portico and will next audit the Scholar's Portal and Hathi Trust)
- spoken with funders about print archiving projects
- convened a group of Provosts of major universities (on CRL's Board) to explore arguments for collective approaches to print preservation.

Lizanne Payne presented CRL's proposal for a collaborative North American print archiving program using slides based on the document she had distributed by email in advance of the meeting (1/12/10). The discussion document and presentation slides are available on CRL's Print Archives website at:

http://www.crl.edu/archiving-preservation/print-archives/action-agenda

She then led the group in discussion of CRL's proposal. The group generally endorsed its provisions and complimented CRL on its clarity and ambition. Several consortia expressed interest in working with CRL on the steps outlined in the document, and other interested groups should contact Lizanne (lpayne@crl.edu). It is likely that a meeting of interested parties will take place in Chicago in April in tandem with ICOLC.

Two topics generated considerable discussion: a) inclusion of academic organizations in overall preservation planning in order to help gain traction with faculty, and b) developing marketing and education materials about print preservation programs, perhaps on the model of SPARC's "create change," for use in campus discussion. There was widespread agreement that these could be important elements of the library community's strategy for building support on campuses.

C. List of participants:

Bernie Reilly (CRL)
Bob Kieft (Occidental College)
Kimberly Armstrong (Center for Library Initiatives, CIC)
Judith Russell (University of Florida)
Martha Brogan (University of Pennsylvania)
Kimberly Douglas (Caltech)
Julia Gammon (University of Akron)
Michael Levine-Clark (University of Denver)
Lizanne Payne (Consultant, CRL)
Kathleen Richman (Law Library Microform Consortium)
Mark Sandler (Center for Library Initiatives, CIC)
Roger Schonfeld (Ithaka)
Michael Spinella (JSTOR)
Melissa Trevvett (CRL)
Catherine C Wilt (LYRASIS)
Millie Jackson (University of Alabama)
Dan Iddings (PALCI)
Karla Hahn (ARL)
Deb Rollins (University of Maine)
Chris Loring (Smith College)
Jay Schafer (University of Massachusetts, Amherst)
Peggy Seiden (Swarthmore College)
Joni Blake (GWLA)
Dennis Massie (OCLC Research/RLG Programs)
Ellen Dunlap (American Antiquarian Society)
Tim Cherubini (LYRASIS)
Michael Stoller (New York University, absent)
Susan Singleton (CARLI)
Dalia Corkrum (Whitman College)
Kathy Scardellato (University of Toronto)

Dustin Holland (Better World Books)
Rick Lugg (R2 Consulting)