Cataloging Hidden Special Collections and Archives:
Building a New Research Environment

A Proposal to the Andrew W. Mellon Foundation
Submitted by the
Council on Library and Information Resources

> By recovering things from the past or by looking at the experience differently, we can see how to think and to act differently in the future. The past can challenge us with eloquent, brilliant, troubling material that widens our present experience and wisdom. It provides perspectives to engage, accounts to cross-examine, and opportunities to hone skills of empathy, compassion, and reflection.
>
> Roy Rosenzweig (with David Thelen), *The Presence of the Past: Popular Uses of History in American Life.*

Overview and Scope of the Problem

This proposal addresses a daunting, pervasive problem that has been discussed and researched for nearly ten years: that our libraries, archives, and cultural institutions "collectively hold millions of items that have never been adequately described and therefore are all but unknown to, and unused by, the scholars it is our mission to serve" (Tabb, 123). The phenomenon was initially described in some detail in a 1998 ARL survey of 99 North American research universities' special collections. That report found that on average 15 percent of printed volumes in special collections remained unprocessed or uncataloged, and thus hidden. The figure rose to an average of 27 percent of manuscripts, and 35 percent and 37 percent for video and audio respectively. Nationally, this represents a staggering volume of items of potentially substantive intellectual value that are unknown and inaccessible to scholars, causing some library directors to refer to the problem as a "dirty little secret" (Tabb, 123) or "the elephant in the closet" (Mandel, 106).

In addition to the unanimity accorded to the seriousness of this problem, the reports, proceedings, and white papers devoted to the topic also share a common theme pertaining to its solution: that a national effort must be undertaken to coordinate the cataloging of these rare and hidden materials, and that the records and descriptions obtained through such an effort must be accessible through the internet and the Web; allocating funds to produce only locally obtainable records would fail to take advantage of the digital environment that would expose these collections to a global audience of scholars, students, and teachers as well as facilitate the harvesting, aggregation, and thematic correlation of the records to advance intellectual productivity.

The more recent reports on hidden collections also emphasize the need to make the categories and schemes of record creation and descriptions less rigid than past practice, as the tightly defined fields are increasingly seen as impediments to a more interoperable

accumulation of pertinent data. This response represents a profound change of mind-set from the past: cataloging special collections and archival materials has routinely been defined as a local practice. The shift to understanding the hidden collections problem as a national responsibility entails an acknowledgement that for the 21st century collaboration, coordination, and coherence of response of cultural institutions to the community of users is fundamental and takes precedence over local practice. As succinctly stated in the recent report from the Library of Congress on the future of bibliographic control, "different communities of bibliographic practices have grown up around different resource types: library collections of books and journals, archives, journal articles, and museum objects and images. As these resources and others become increasingly accessible through the Web, separation of the communities of practice that manage them is no longer desirable, sustainable, or functional" (LC, 7).

Outline of Proposal

The Council on Library and Information Resources (CLIR) proposes to organize and instantiate a program to identify and catalog currently hidden special collections and archives. This proposal incorporates the priorities articulated in the background research and recommendations on the phenomenon of hidden collections: it is national in scope; the methodologies adopted by the funded projects need to be broadly applicable and subsequently built upon over time; and the means of data creation must be cost effective, efficient, and assure that a critical mass of trusted and authoritative information is achieved quickly. CLIR's strategy for building this distributed organization assumes local autonomy and responsibility but which also requires centralized agreements concerning governing principles that will ensure enterprise-wide coherence. In this way the organization structure mirrors the technical implementations.

We request funding for one year, in the hope that, with sufficient response to the first round of proposals and subsequent execution of the individual funded projects, additional grant requests over a period of five years would be considered. Grants to institutions or consortia of institutions will have a three-year term; CLIR will write a report at the conclusion of the three year period, as well as submit annual updated reports beginning in 2009. The Review Panel will be formed immediately following the announcement of the grant, with the RFP posted in the spring of 2008 and a deadline for proposals set early in the fall of 2008.

In all aspects of the program CLIR will adopt a lean and flexible approach. The scale and scope of this program will entail considerable influence and likely encourage standardization and efficiencies that a more narrowly conceived effort might not. The onus, however, to achieve the desired goals of interoperability, effective harvesting, and longer term sustainability will fall to the institutions as requisite for individual project funding. In this respect the main structuring device of the program is the Review Panel and its salient instrument, the Request for Proposals (RFP). There will be a wealth of expertise and knowledge that will be brought to bear on a variety of facets of this program (platform adoption, digital architecture, protocols, knowledge organization) by means of advisory committees. The advisory committees will routinely report to the Review Panel, but will also work directly with institutions participating in the program.

In this regard the universities, colleges, museums, and other cultural institutions awarded funding for cataloging hidden collections will benefit directly through the exchange of reports and email, and occasional teleconference calls, with representatives from the pool of experts to better coordinate and bring coherence to their efforts as well as to assure persistence of the information generated by means of a standardized, interoperable approach.

Institutions of higher education and cultural organizations that hold important collections that are difficult or impossible to locate through finding aids will submit proposals in response to the posting of the RFP. The RFP will include a set of data fields that require information from the applicant pertaining to the scope and depth of the hidden collection, its disciplinary focus, value to research, type of media, and other descriptive elements that will assist the Review Panel in assessing the intellectual impact of cataloging and making visible these materials. The RFP will also require responses from the applicant about long-term sustainability, additional sources of funding, and evidence of institutional support. The rationale for this program and its aspirations will also be detailed in a prologue to the RFP. The Review Panel will scrutinize the submitted proposals, and will call upon field experts when necessary. Scholars will determine the final selection of applicants to receive awards, always with the aim of significantly improving access to materials of fundamental importance for research and teaching as adjudicated by those expert in the appropriate fields of study.

All non-confidential information gathered and generated through this program (sections of the RFP from applicant institutions, and the catalog records of the funded projects) will be available via the Web, exploiting the power of the Internet to federate disparate, local cataloging entries with tools to aggregate this information by topic and theme, and with a concomitant goal of stimulating research and providing opportunities for new questions and methodological approaches. By exposing collections, it is hoped that the cataloging phase over time will instigate subsequent programs that will digitize some or most of the cataloged materials. This proposal thus takes a cyberinfrastructure model as the basis of its design: each phase of gathering information will inform the next, with the whole conducing to a new environment in support of scholarly productivity. The longer term aspirations would reflect the more encompassing definition of bibliographic control espoused by the Library of Congress in its recent report: multiple kinds of information, including catalogs, descriptions, notes, syllabi, and other aspects of scholarly communication would be unified in a rich environment capable of an unprecedented depth and breadth of discovery.

A Note on Cyberinfrastructure

This program provides an opportunity to construct a new research and teaching environment of national importance and, in keeping with a key tenet of cyberinfrastructure, to facilitate building new communities of interest and research that are simultaneously locally grounded and independent but cooperative across the entire enterprise. These communities are defined as virtual organizations that transcend geographic and institutional boundaries, an interlocking of technical and social elements.

Three layers of information can be artificially defined (artificial in the sense that the information strata will be interoperable to the user and machines): (1) a basic registry of hidden collections and archives that can be found through a Web-based platform; (2) a descriptive record of a subset of collections that are deemed most urgently in need of cataloging and documentation; and (3) digital versions of the special collections and archives that have been cataloged. This program focuses on the first two layers but will be developed with an eye to the eventual development of the third, subject to future funding. Over time, new tools and applications can be built to take advantage of the large datasets, which can support new questions and methodological approaches.

The first layer—the basic registry—will be created from information in the proposals submitted. Applicant institutions will be required to agree that the information they submit will become part of this registry, whether the collections are cataloged or not. The rough, initial posting of the titles, subject areas, and locations of collections should make future cataloging easier, as funding can be better directed by the Review Panel to disparate but thematically coherent collections that might not otherwise be exposed. The registry will from the outset be coordinated with other programs such as ArchivesUSA and the University of Idaho's Repositories of Primary Sources.

ArchivesUSA is a current directory of 5,581 repositories and 160,792 collections of primary source material across the United States. Each collection record links to its corresponding repository record. Collections include the National Union Catalog of Manuscript Collections (NUCMC) from 1959 to the present, covering more than 106,000 collections, and names and detailed subject indexing of over 64,000 collections whose finding aids have been published separately in ProQuest UMI's microfiche series, National Inventory of Documentary Sources in the United States (NIDS). Idaho's list of Repositories of Primary Sources is a collection of links to web sites that describe physical collections of rare books, manuscripts, archives, historical photographs, oral histories, or other primary sources. The list focuses on actual repositories and does not include virtual collections or transitory exhibitions. Given the ease of registration, the applicant institutions will be requested to link their proposal information with the appropriate national database.

The second layer—a federated digital catalog—will evolve as the funded proposals are completed. Funding of proposals will be contingent upon applicants describing how they will make their catalogs available and integrated with other catalogs of revealed special collections and archives. Over time, it, too, will become a trusted national resource. The scale of this program should ensure more standardized approaches to the cataloging of special collections and archives, making it easier for collaborative efforts to contribute to the enrichment of the data. The desired third layer—a federated digital library of special materials—will bring the original sources even closer to students and scholars.

As in other aspects of cyberinfrastructure development, the implementation of this program will require the collaboration of librarians, archivists, scholars, and technical

experts. Every effort will be made to minimize the costs and time of program administration, and to exploit the power of the Internet and Web for data gathering and information organization. This program is explicitly about effecting a major change of practice pertaining to the creation, maintenance, and disclosure of records of hidden collections.

In the end, a fundamental facet of our cultural commonwealth will be revealed, organized, made accessible, and preserved. The agenda of the Council on Library and Information Resources fits very well with the design and outcomes of this program; CLIR is committed to both inculcate the importance and implications of a cyberinfrastructure-based design, and contribute to the instantiation of cyberinfrastructure environments whenever possible.

Specific Elements of the Program for Cataloging Hidden Special Collections and Archives: Building a New Research Environment.

This section is organized by the following subjects: The Request for Proposals (RFP): General Assumptions and Data Fields; Program Organization; Grants; Program Timeline and Activities; and Conclusion.

The Request for Proposals (RFP)

The RFP is the main structuring instrument of this program. Several questions, raised at the outset of discussions, have identified criteria for eligibility, and general project requirements. In Part One, some of these assumptions are articulated. Part Two lists some of the key fields of information that the RFP will solicit. These assumptions and questions that will structure the RFP are based on extensive research on the problem of hidden collections and wide consultation. Members of the Review Panel will be selected in part in response to their affirmation of these general principles and methodology. Additional areas of focus or modifications to these assumptions and the RFP's formulation will be subject to ongoing discussion, especially as the program evolves; the parameters articulated in the RFP sections are nonetheless presented with a very high degree of confidence and a correlate commitment to maintain them.

Part One: General Assumptions

Rationale for the Program

This grant program addresses the pervasive problem of the inability to know about and access a significant percentage of special collections and archives that are of considerable value to research and teaching.

Conceptual Approach

The process for revealing the hidden collections will involve adopting a technology platform (or platforms) that will allow for accurate descriptive information to be entered quickly, efficiently, and cost effectively. Graduate students and paraprofessionals will be

trained to do most of the data entry. Each funded project will be linked to and interoperable with all other projects funded by this grant, to form a federated environment that can be built upon over time. Institutions must acknowledge local ownership of the data generated through this award program, and agree to its persistence.

Technology to be Used

No one technological platform will be designated. Technical solutions exist (see Appendix 2) for consideration, that allow for swift and efficient data entry that is then translated into standard records format such as EAD and MARC. Institutions and consortia awarded grants will work together to find the most effective approaches that meet the overall requirements of the program.

Existing models that can be adopted for the execution of this program

There are several aspects of ongoing projects that could be appropriated to save time and expense. One promising example is the Chicago model of cataloging hidden collections developed by Prof. Jacqueline Goldsby. This model relies on graduate students to catalog those collections discovered through the process, in this case, of a national solicitation to reveal hidden yet valuable materials. The graduate students work with a well-defined template of descriptive data fields that does not require a high level of cataloging expertise. The graduate students are not fellows but salaried employees of their respective institutions; all work is done in rigorous consultation with faculty and librarians. The graduate students bring a disciplinary expertise to the projects, and learn firsthand the challenges of organizing information, creating digital records, and other fundamental issues of librarianship. Because the students are often working with collections in their area of study, they are engaged with source materials that may enhance their mastery of scholarship. See http://www.arl.org/bm~doc/arlbr251uncap.pdf for a concise description of this project and its goals.

Another is work being done at the University of Illinois, Urbana Champaign (UIUC). This is also a hidden collections project with a very successful track record. It is a smaller scale version of the University of Chicago effort, and generally follows the structure and staffing of the latter. The technological component of the UIUC project is well documented and noted in the Readings section.

Grants will be allocated to institutions or consortia that agree to employ graduate students, paraprofessionals, and other staff that will contribute to a cost effective and swift generation of records.

Benchmarks

Insight to the celerity and accuracy of the models noted above can be gleaned by the statistics from UIUC, which has cataloged 20,000 books from its special collections in 14 months.

Funding levels of project grants
An array of grants would probably better address the complexity of this problem. Each grant will have a term of three years. Because this first round represents both an inauguration of a new approach to the hidden collections problem and concomitantly a test some basic assumptions, the grants may tend to range from $100k to $500k. The actual outlay of funds will nonetheless be determined by the annual context the proposals themselves create.

Definition of terms
"Special collections" and "archives" are established terms and, while subject to slightly different institutional interpretations, are serviceable. We will take *special collections* to mean rare, often-unique materials generally housed in secure, monitored environments. *Archives* can pertain to unique collections associated specifically with an individual or organization. By *not* defining these terms prescriptively, we hope to encourage a process that is more encompassing and revelatory by allowing broad interpretation among respondents.

Criteria for selection of submitted requests for cataloging of hidden collections
The main criterion for determining the priority of the collections to be cataloged is the value of the materials for scholars and students. Rather than solicit for collections by a specific topic, the steering committee will see what kinds of collections are submitted by the various institutions and consortia and set priorities on the basis of that pool. Scholars increasingly work in a digital environment and are interested in finding related collections across many institutions. Consequently, collaborative proposals that aggregate disparately located but similarly themed collections will be more favorably weighed. Alternately, the steering committee might aggregate several candidate collections as a single project.

Conservation as an element of grant consideration

The physical condition of the materials will be considered, but will not be a determining factor initially. The focus of this program is not conservation and no grants will be allocated for conservation purposes. Similarly, funds will not be allocated for retrospective conversion.

Existing finding aids

The collections that will be candidates for cataloging are by definition hidden. Some, though, may have some form of finding aid, but provide no value to scholars. If there are methods to reconstitute some of the finding aids into a unified platform that is cost effective, this should be considered in the grant program.

Formats

The range of media that can be termed special collections or archives should not be restricted. Increasingly, valuable collections are composed of many formats: paper,

moving images, sound recordings, postcards, photographs, and other realia. Since one goal of the program is to assess the scale and scope of hidden collections, a broad definition of what constitutes an appropriate artifact or medium will be used.

Geographic boundaries of the program

The focus of this program will be on collections owned or held in U.S. institutions. Any opportunity, however, to collaborate with similar cataloging efforts overseas will be pursued to broaden the adoption of agreed-upon standards for access and aggregation.

Part Two. Data Fields in the RFP

The electronic RFP will require some version of each of these data fields; the form itself will be divided into sections that ask for information that will become public, as part of the registry, and information that will remain confidential to the Review Panel. The public information can in this way be more easily extracted from the form and ingested into the database.

- Are the collections of significantly high value to scholars? In what ways? Will scholarship be enriched by cataloguing and making accessible these materials?
- Are there other collections that pertain to the theme or content that could be aggregated to produce a more coherent and encompassing set of records?
- Is the project plan reasonable and deliverable? How will the proposed quantity and type of records described be catalogued in the time specified?
- How does the institution insure that the records generated by this project will be interoperable, efficiently harvested, and sustained? By what means does the institution insure that its data will be federated with other participating organizations?
- What are some examples of past collaboration that resulted in shared data or federated information programs?
- Has the institution or organization set priorities for its cataloging backlog and does this project address one or more of the priorities identified?
- Does the institution hold the rights to make the collections to be cataloged, and can allow these collections to be accessed and, later, digitized?
- What infrastructure or other elements of technical support does the institution require that would be funded by the grant?
- How does the institution propose to staff the project? What kinds of training would be requisite to assure proper execution of the proposal?
- Does the institution have a strategic approach to addressing its cataloging backlog?
- How will the institution ensure access to the collections once they are revealed?
- Does the project initiate or support other desirable future developments?
- Does the collection have conservation needs that would limit the impact or feasibility of cataloging work?
- Is it reasonable for this organization to seek external funding for cataloging this collection?

- Does the applicant agree to assist in the evaluation of the project following completion?
- What institutional or local expertise can be brought in to the project?

Each proposal will also need to include a general timeline, with deadlines, goals, and benchmarks indicated, for the project.

Program Organization

The *Review Panel* is the committee that provides structure, guidance, and oversight to this program, and makes final decisions on the award of project grants. The panel will be composed of ten members:

5 distinguished scholars chosen from different fields of study
1 University library director
1 College library director
1 IT expert with experience in large, nationally coordinated projects
1 Special collections expert
1 President of CLIR

The Panel will be co-chaired by the president of CLIR and a scholar. The composition of this committee is meant to ensure the program's priority of making collections available that are of the highest value to research and teaching. Because of the scope and potential influence this program may entail, representatives from national organizations and funding agencies will be invited to Panel meetings as guests. Staff and associates of CLIR who will also be involved with this program include the Presidential Distinguished Scholar; the Director of Programs; and the Director of Communications. The Review Panel will meet twice a year, and a structured means of ongoing electronic communication will be established at the outset.

Advisory committees will also be established prior to the launch of this program, and will include some of the most accomplished individuals in their respective fields (see Appendix 1 for a list of candidates). These committees will be virtual for the most part, and will be called upon for issues pertaining to their fields of expertise (technology, archives and special collections, issues pertaining to research libraries, museums, and other cultural organizations). Small groups of experts may be convened at CLIR from time to time to discuss especially complex issues should they occur.

Most critically, a network of scholars will be identified who can assist in determining the intrinsic intellectual value of the collections submitted for consideration through the RFP. Because the Review Panel will be limited to five senior scholars, it is not possible to adequately cover the intellectual breadth of the institutional collections this program will attract. ACLS and the Mellon Foundation have agreed to assist in identifying these scholars.

Additional support for this program will be instantiated in the following roles:

Project Manager. Duties would include coordinating the program, assuring financial and other reports are executed on time, and budgets adhered to. The project manager would also serve as the program liaison to the Review Panel and other less formal committees and groups, as well as documenting the program as it evolves. Not the least, we see this position dedicated to working closely with the grantees to assure coherence of mission and interoperability of production data. This would entail ongoing electronic communication and site visits as needed.

Post-doc. An engaged and talented post-doc could bring some disciplinary expertise to the project as well as a current understanding of technology and distributed organizations. The Post-doc would work with CLIR's other post-doc programs, and participate in and closely observe this program from a young scholarly perspective. S/he will give presentations at various information and academic conferences, and write assessments, historical commentary, and conduct research on and about the program over time. Since many of those working on the project will be graduate students, the post-doc provides the perspective of someone familiar with graduate students and a means by which to build a more permanent community of students who are either participating or interested in this program.

Intern. The intern would be drawn from a local university to assist in logistics and execution of some of the chores associated with meetings, travel, and teleconferencing. The intern would work with both the project manager and post doc when needed.

Grants

The premise of this proposal rests on establishing a method of cataloging hidden collections that is extremely cost effective, efficient, and swiftly executed. Put bluntly, money in the form of grants will go a long way. A $75,000 grant over three years to one institution should contribute significantly to the national effort. It is also expected that many proposals will be collaborative efforts by several institutions, suggesting that grants on the order of $250,000 to $500,000 by the larger coalitions of applicants are to be expected. It is difficult to conceive of grants above $750,000, especially in light of the enormous intra-consortial effort needed to manage a project of this scale, and the wiser course may be to focus on grants up to the $500,000 range for the first round to better assess issues pertaining to scholarly impact weighed against complexity of execution.

As the granting agent, CLIR will not award funding for overhead or indirect costs. An annual report from all institutions or consortia awarded grants will be requisite

Program Timeline, Annual Reports, and Activities

The term of this grant proposal is four years, April 1, 2008 through March 2012. Awards will have a three year term; each award recipient will be required to submit to the Review Panel an annual status report, so that grantees will submit to CLIR an annual report in December 2009, 2010, and 2011. CLIR will submit to the Mellon Foundation an annual summary report on the overall program for each of these years, based on the reports

received from the grantees, as well as a report in the fourth year, in March 2012, that summarizes the overall degree of achievement and consequence of this program.

Occasional white papers and research reports on the efficacy of the program and its implications for scholarship and teaching will also be issued; other agencies, such as Ithaka, OCLC/RLG, ARL, and ACRL may also conduct reports about and research on the program from the perspective of national impact (e.g., increased accessibility, evidence of greater collaborative and cross-disciplinary research, emerging methodologies) as it evolves.

April 2008: The Review Panel will be named and convened. All potential candidates for the Review Panel will be contacted and informed of the possibility of the grant prior to March; details of the program and a draft of the RFP will be shared with each potential panel member so that the committee can move quickly to post the RFP.

April 2008: A web page will be designed for the project, to include all relevant information and goals of the project, a general calendar for publications and assessments, a Q&A section, and a draft of the RFP. RFPs will be electronic and the information gathered from their submission will be migrated into a database.

Late April 2008: The RFP will be promulgated, inaugurating the program. Deadline for receipt of proposals will be June 2008.

August 2008: Review Panel convened to review proposals and select projects to award.

Early September 2008: Project awards will be announced.

September 2008-March 2009: routine updates on the program will be delivered at a variety of venues; occasional reports issued; with ongoing informal evaluation and assessment of the program.

Conclusion

The impact and influence of this program could be profound. On one level, it will make visible and accessible valuable collections that would otherwise remain out of reach to scholars and students. By revealing these materials, the program may instigate new collaborations among scholars and facilitate new approaches to research: new questions may be asked, and new kinds of queries can be made against the data. In response to its scope, there may be greater impetus to adopt its technical platform, which in turn may lead to an acceptance of standards that is truly national, something that has been difficult to achieve in the last decade. Existing and future cataloging efforts would build upon this program more cost effectively, enriching the database through federation and shared services.

The cyberinfrastructure design that permeates this program is by nature extensible; multiple layers of applications and tools can be built over time, as well as ongoing

integration of the digital representations of the materials cataloged. In this respect the program to catalog hidden collections is more a facet of a longer term process that bridges communities of interest, allows for collaborative sharing at an unprecedented scale, and contributes to a robust, thoughtfully managed and sustained intellectual commonwealth.

CLIR/
cjh
01/04

Appendix 1: Project Organization

At present, a number of highly qualified individuals have expressed interest in this proposal. A Review Panel, pivotal to the successful execution of the project will be convened if the grant is awarded, and composed of scholars, library directors, and specialists in the fields of special collections and archives. Other advisory groups will be brought together informally (generally by email or conference call) as issues are raised and new directions identified. Those who have expressed interest are listed below, with some groupings of expertise suggested, as well as some names who have been suggested as potential program advisors.

Scholars
Asif Agha, University of Pennsylvania
Susan Alcock, University of Michigan
Richard Bulliet, Columbia University
Terry Castle, Stanford University
Kathleen Conzen, University of Chicago
Jacqueline Goldsby, University of Chicago
Tony Grafton, Princeton University
Hans Hillerbrand, Duke University
David Der-wei Wang, Harvard University
Steve Wheatley, ACLS
Christoph Wolff, Harvard University

Librarians
Paula Kaufman, University of Illinois, Urbana-Champaign
Michael Keller, Stanford University
Anne Kenney, Cornell University
Rick Luce, Emory University
Carol Mandel, New York University
Deanna Marcum, The Library of Congress
Alice Prochaska, Yale University
Elliott Shore, Bryn Mawr College
Suzanne Thorin, Syracuse University

Special Collections/Archives
Sid Berger, Philips Library of the Peabody-Essex Library in Salem, MA
Christopher de Hamel, Librarian of Corpus Christi Cambridge
David Gilner, Hebrew Union College
Mark Green, SAA
Eric Holzenberger, Grolier Club
Valerie Hotchkiss, University of Illinois at Urbana-Champaign
Deborah Leslie, Folger Shakespeare Library
Paul Needham, Princeton
Richard Oram, Harry Ransom Humanities Research Center, Austin (U.T.)
Margaret Powell, Lewis Walpole Library, Yale

Ben Primer, Princeton
William Stoneman, The Houghton, Harvard University
Richard Szary, University of North Carolina at Chapel Hill
Daniel Traister, University of Pennsylvania

Advisory Committee for Information Technology
Sayeed Choudhury, Johns Hopkins University
Geneva Henry, Rice University
Katherine Kott, Aquifer
Chris Prom, University of Illinois, Urbana/Champaign
Ron Larsen, University of Pittsburgh

Senior Consultants
Stanley N. Katz, Princeton University
Christine Borgman, UCLA
Paul Courant, University of Michigan

Steering Committee for National Impact, Project Adoption, and Sustainability
Allied Organizations Committee
Presidents/Executive Directors of:
ARL
CNI
CLIR
CRL
DLF
Ithaka Harbors
SPARC
RLG/OCLC

ARL Point of Contact
Chair, Working Group on Special Collections

Funding Agencies Liaisons
IMLS: Joyce Ray
NEH: Joel Wurl, Brett Bobley, Suzanne Lodato
NSF: Lucy Nowell, Sylvia Spengler, Steve Griffin
NCO: Chris Greer

Appendix 2: Technology Platforms

One platform that might be considered as a model to adopt is the **Archon** program at the University of Illinois, Urbana-Champaign. Archon is being used for a cataloging program focusing on hidden collections at the local level, and has proven effective. It is a unified web-based platform for archival description and access, built on a LAMP software stack (current versions of Linux, Apache, MySQL, PHP). It can be used by graduate students with proper supervision, since no knowledge of data structure standards (such as MARC or EAD), or even of Archon's internal database structure the structure, is needed to list and catalog collections. Archon automatically generates an EAD and collection-level MARC record, and can link to external descriptive records, such as box lists stored in PDF files or other formats. Archon also includes a templating system, so that additional outputs can be defined with relatively little effort. The platform generalizes functions to a very high degree, which contributes to a fast and efficient search result.

Among its enhancements, Archon allows scholars and students to:

- simultaneously search for relevant archival collections, series, files and items, as well as associated digital objects;
- easily navigate from digital objects to associated collection descriptions and vice versa;
- view search hits in their full archival context;
- browse materials by collection title, digital object title, controlled subject heading, creator authority record, or archival record group;
- jump easily between collections and digital objects sharing the same subject, creator, or archival record group;
- show finding aids in print view; and
- view or download digital objects.

Program staff can:

- create standards-compliant collection descriptions and full finding aids using Web forms;
- log into an enhanced 'staff view' of the public website;
- move seamlessly from the public view to the editing interface;
- describe the series, subseries, files, items, etc. within each collection;
- organize collections into record groups and subgroups based on provenance or any another scheme defined by a repository;
- develop creator authorities and link them to collections and digital objects;
- manage a controlled subject list and link subject terms to collections and digital objects;
- upload digital objects directly into archon or link archival descriptions to external URLs;
- link digital objects to their parent collection, series, or file;
- login to an enhanced public interface; and
- export MARC and EAD records (for importation to other systems).

Another platform that could be adopted for this project is the **Archivists' Toolkit**. Some of its key features include:
Key Features:

- Integrated support for managing archival materials from acquisition through processing:
- Recording repository information
- Tracking sources / donors
- Recording accessions
- Basic authority control for names and topical subjects
- Describing archival resources and digital objects
- Managing location information

Customizable interface:

- Modify field labels
- Establish default values for fields and notes where boilerplate text is use
- Customize searchable fields and record browse lists
- Ingest of legacy data in multiple formats: EAD 2002, MARC XML, and tab delimited accession data
- Rapid data entry interface for creating container lists quickly
- Management of user accounts, with a range of permission levels to control access to data
- Tracking of database records, including username and date of record creation and most recent edit
- Generation of over 30 different administrative and descriptive reports, such as acquisition statistics, accession records, shelf lists, subject guides, etc.
- Export EAD 2002, MARC XML, METS, MODS, and Dublin Core
- Support for desktop or networked, single- or multi-repository installations

Appendix 3: Further Opportunities for Partnerships

One of the most compelling facets of this grant proposal is its potential to form substantive and last collaborations within the larger community of libraries, cultural organizations, and higher education at large. Related work in **Ithaka** on Aluka records and work on the Archivists' Toolkit are natural points of convergence. This program will also be closely with the **Office of Cyberinfrastructure** and other relevant agencies. While the content of the hidden collections will most likely be of more immediate value to the humanities and social sciences, the digital environment that this program builds is ultimately discipline-agnostic.

As of this writing, **NEH** has expressed interest in this program as a potential additional funder. NEH and other agencies routinely fund cataloging programs; ideally, CLIR would establish a program infrastructure that would allow for these agencies to invest in cataloging hidden collections while avoiding the overhead costs of constituting their own external review panels and other redundant aspects of awarding similar grants. **IMLS** is focused more on conservation, so one avenue of approach could be to work closely with that agency and the programs it funds, extrapolate data associated with those conservation efforts, and propose that cataloging follow as a logical next step. Often, the collections that benefit from the conservation grants are hidden, so the IMLS program can be seen as a means to reveal the valuable materials, conserve them, and integrate them through cataloging and other finding aids into a national database by way of this program.

A formal, programmatic collaboration with **ARL** has been discussed, and will be instantiated if the proposal is funded. ARL has long be interested in, and actively researching, the issue of hidden special collections. The chair of the Working Group on Special Collections will be the first line representative from ARL to this program. The working group is embedded in ARL's initiative on Research, Learning, and teaching, which is particularly apt in light of the goals of this proposal.

Another potential partner over the longer term is **DLF Aquifer**. Although this program will not digitize selected collections, it is essential to design procedures for cataloging hidden collections, and the resultant records, so they can eventually be scanned for sharing in a federated digital library. To assist in this design, representatives of DLF's Aquifer should be incorporated into the planning and execution of the cataloging program from the start. Aquifer was created in part as a response to the proliferation of digital collections that are difficult to use and even more difficult to find. It may be prudent to similarly frame the cataloging of hidden collections within the mission of Aquifer as an example of a nascent effort that is constructed with the persistent access and utility of digital resources for scholars as its paramount aim, and concomitantly explore developing Aquifer as a repository of tools and analytic applications. A national digital library will consist of content and tools; if Aquifer focuses on tool identification, evaluation, and availability, it might fundamentally complement the various digitization projects underway.

Selected Readings

ARL Research, Teaching, and Learning. Exposing Hidden Collections Conference Summary, 2003. http://www.arl.org/rtl/speccoll/hidden/EHC_conference_summary.shtml

ARL Research, Teaching, and Learning. Task Force meeting with the Archival Community. October 2004. Summary at http://www.arl.org/rtl/speccoll/spcolled/sptfmins1004.shtml.

Atkins, D. et al. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure.* Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. http://www.nsf.gov/od/oci/reports/toc.jsp

Baker, S., and Millerand, F. 2007. *Scientific Infrastructure Design: Information Environments and Knowledge Provinces.* Proceedings of the American Society of Information Science and Technology (forthcoming).

Brogan, M. 2006. *Contexts and Contributions: Building the Distributed Library*. Digital Library Federation. http://www.purl.oclc.org/DLF/pubs/dlf106/

Edwards, P.N., Jackson, S.J., Bowker, G.C., and Knobel, C. 2007. *Understanding Infrastructure: Dynamics, Tension, and Design*. NSF Report of a Workshop: History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. http://hdl.handle.net/2027.42/49353.

Erway, R., and Schaffner, J. 2007. Shifting Gears: Gearing up to get in the Flow. OCLC Programs and Research Publication. www.oclc.org/programs/publications/reports/2007-02.pdf

Greene, M. and Meissner, D. 2005. More Product, Less Process: Revamping Traditional Archival Processing. *American Archivist* 68: 2, 213.

Heritage Preservation. 2005. *A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections*. Heritage Preservation and Institute for Museum and Library Services.

Jones, Barbara M. "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in America's Research Libraries." RBM vol. 5, no.2. Fall 2004. 88-105.

Lack, R. 2006. The Importance of User-Centered Design: Exploring Findings and Methods. *Journal of Archival Organization* 4: 71-78.

Mandel, Carol. “Hidden Collections: The Elephant in the Closet.” RBM vol.5, no.2. Fall 2004. 106-113.

NSF Cyberinrastructure Council. 2007. *Cyberinfrastructure Vision for 21st Century Discovery*. http://www.nsf.gov/od/oci/CI_Vision_March07.pdf.

Panitch, Judith. Special Collections in ARL Libraries. Results of the 1998 Survey Sponsored by the ARL Research Collections Committee. (Washington, D.C. : ARL, 2000)

Prom, Christopher J., Christopher A. Rishel, Scott W. Schwartz, and Kyle J. Fox. 2007. “A Unified Platform for Archival Description and Access.” *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries*, 157-166. Available: http://doi.acm.org/10.1145/1255175.1255205

Prom, Christopher J., forthcoming, 2008. “Optimum Access? A Survey of Processing in College and University Archives.” in *College and University Archives: Readings in Theory and Practice*. Chicago: Society of American Archivists Press. 53 pp mss. Preprint available: http://web.library.uiuc.edu/ahx/workpap/ChapterEight-Prom.pdf

Report on the Future of Bibliographic Control. Library of Congress Working Group on the Future of Bibliographic Control. Review Draft. http://www.loc.gov/bibliographic-future/

Rosenzweig, R. et al. (2006). Our Cultural Commonwealth. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure in the Humanities and Social Sciences. http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf

Szary, R. (2005) Encoded Archival Context (EAC) and Archival Description: Rationale and Background. *Journal of Archival Organization* 3: 217-227.

Tabb, Winston. “’Wherefore Are These Things Hidden’? A Report of a Survey Undertaken by the ARL Special Collections Task Force.” RBM vol. 5., no.2. Fall 2004. 123-126.