Digitizing Hidden Special Collections and Archives
Enabling New Scholarship through Increasing Access to Unique Materials

Program Proposal
Submitted by the Council on Library and Information Resources
to
The Andrew W. Mellon Foundation
September 17, 2014

Section I. Background

For seven years, The Andrew W. Mellon Foundation has made significant investments in innovative and efficient approaches to describing rare collections through *Cataloging Hidden Special Collections and Archives: Building a New Research Environment*. The impetus for this program was a widely shared acknowledgement of the need to rethink cataloging methods toward greater standardization and efficiency in the interest of supporting scholarship and teaching. The urgency of this need, explored through a decade of research beginning in the 1990s, compelled the Foundation and the Council on Library and Information Resources (CLIR) to create a national program to address it, a program that would fund locally produced records that would be available through the internet and the Web.

Cataloging Hidden Special Collections and Archives was never intended to be a long standing program. It was conceived as a kind of incubator which, if successful, would inculcate an

understanding within the library and archival professions that while all special materials are by nature local, creating standardized descriptions of them that can be accessible anywhere and anytime is an exceptionally important goal. One of the more critical of the desired outcomes of the program was a change in professional behavior that would remain active after the program concluded, and concomitantly a willingness to integrate more efficient and effective approaches to cataloging into the normal operations of individual institutions. Through encouraging grant recipients to share their experiences with others, CLIR and the Foundation also hoped to see the perpetuation of these new approaches across institutions and professional communities.

As an incubator, the program has succeeded in meeting its aspirations. Projects funded through it have allowed for large quantities of materials of integral value to scholarship to be exposed and discoverable. Through a variety of presentations, publications, and social media,[1] grant recipients have shared their approaches, and these have proved useful models for others. In light of this wide adoption of creative and efficient techniques, we believe that library and archival communities are now willing and able to continue revealing materials of rare and special scholarly value framed by the techniques and habits the program inaugurated, and will also seek local funding and external grants to continue to do so.

Despite the fact that large amounts of materials still await cataloging in cultural heritage institutions, and despite CLIR's recognition that this work is still important, we believe it is not necessary to continue funding *Cataloging of Hidden Special Collections and Archives* in its current form as a program of CLIR. In keeping with CLIR's aim to inculcate positive changes in practice that support the creation of new knowledge, we believe that the logical next step involves facilitating *complete* access to rare cultural artifacts online. We remain committed to projects in which locally executed protocols contribute to a national good, using methods that are cost efficient and subject to wider adoption. For this next iteration, we aspire to support the creation of digital representations of unique content of high scholarly significance that will be discoverable and usable as elements of a coherent national collection.

This proposed successor program will enhance the emerging global digital research environment in ways that support new kinds of scholarship for the long term. Its aim will be to ensure that the full wealth of resources held by institutions of cultural memory becomes integrated with the open Web. To promote broad access, careful preservation, standardization, and usability, approaches to digitization should be coordinated across institutions when feasible. By encouraging strategic collaboration and communication among this program's grant recipients, CLIR expects to help broaden understanding of the complexity of these issues in the professional communities responsible for rare and unique collections.

---

[1] See CLIR's website at http://www.clir.org/hiddencollections/resources for a selection of materials provided by grantees to CLIR for dissemination.

Section II. Rationale

To inform decision-making about the potential new program, CLIR and the Foundation have been engaged in an intensive research and consultation process throughout the first half of 2014. CLIR first outlined this process in its January 2014 proposal for a planning grant to support it. The research has included a scan of relevant literature, two consultation meetings with experts in the funding and practitioner communities, and a preliminary review of publicly available data about digitization projects recently funded through national-level programs [See **Appendix 2**]. While this process is still ongoing, CLIR has gathered enough data to articulate five core values that can inform program design moving forward:

Scholarship: The program will be designed to maximize its impact on the creation and dissemination of new knowledge. "Scholarly significance" has been the primary criterion for *Cataloging Hidden Special Collections and Archives* since 2008, and CLIR proposes to keep this as the central requirement for its successor. The program should reveal rare and unique materials that have high significance for scholars' work. A closely related goal is to encourage approaches to digitization that make possible new kinds of scholarship in the digital research environment, including consumption and analysis of content through automated means such as text and data mining.

Comprehensiveness: The program will support the digitization of *entire* collections of significant scholarly value. While selective digitization, or digitization on demand, is a routine practice in many institutions and can be important for increasing access and aiding discovery, the program's investment will be in projects that expose large amounts of related material within a concentrated period of time. Limiting the scope of the program to the digitization of whole collections, or at least large portions of collections, will ensure that applicants justify the significance of all the materials they nominate for digitization, help applicants more accurately specify costs and outcomes, eliminate the need for recipients to invest valuable project time on selection, and maximize the program's potential impact on scholarship.

Collaboration: The program will promote strategic partnerships rather than duplication of capacity and effort. Digitizing unique materials appropriately requires equipment and expertise uniquely suited to the format and condition of those materials. The majority of cultural heritage institutions cannot afford to maintain the capacity to handle and scan the wide array of objects held in their collections, much less maintain the infrastructure for accessing and preserving the digital files they create. For this reason, the formation of strategic partnerships among institutions is critical to establishing cost-effective, sustainable practices for digitization. These partnerships can take a variety of shapes. They can persist for the life of a project or long-term, and may involve commercial as well as not-for-profit entities. They can involve shared equipment, labor, expertise, discovery services, or preservation infrastructures.

Sustainability: The program will promote best practices for ensuring the long-term availability and discoverability of digital files. Beyond encouraging effective partnerships, the

program should fund projects that take reasonable approaches to digital preservation and to aggregation of metadata for maximum discoverability, approaches that reflect an awareness of current best practices and resources in both domains. The goal will be to support cost-effective and replicable approaches that maximize efficiency of throughput and consistency across institutions.

Openness: The program will ensure that digitized content be made available to the public as easily and completely as possible. While recognizing that collecting institutions often do not control the intellectual property rights for their holdings, the program's priority will be to invest in projects that make widely accessible large quantities of content that is of high significance to scholarship,[2] without the imposition of new restrictions on use. At a minimum, metadata created through funded projects should be dedicated to the public domain; in this, the program will encourage compliance with the policies and standards established for the Digital Public Library of America for ensuring broad access. Ideally, digitized content should be made fully accessible for free. The program will recommend that rights statements and attribution requirements be displayed prominently so that users understand any applicable restrictions. In taking this position, CLIR acknowledges that there is widespread need for education on copyright and intellectual property in scholarly and professional communities. While CLIR is unable to take responsibility for providing such education alone, it will endeavor to encourage others to create and disseminate educational resources on the topic of intellectual property as it relates to the digitization of rare and unique materials.

If funded, the Hidden Collections Digitization program will be managed as an element of an emerging, interrelated digital environment that is the purview of the [Committee on Coherence at Scale](#). The encompassing vision of the Committee insists on understanding large scale projects and programs as aspects of a larger whole that is most efficiently built and maintained as a functioning system. The Committee promulgates a set of principles and working assumptions that will help guide this program. Some of those working assumptions include:
1. The current array of digital projects offers a rare opportunity to think about the feasibility of a new, robust digital environment for higher education that, if designed as a system, would create a virtual educational ecology that would correlate many aspects of knowledge organization, the cycle of scholarly communication, knowledge discovery, and pedagogy;
2. A well-designed environment that correlates these various facets of scholarly communication should enhance productivity and encourage new discovery. Working within this multifaceted environment will also foster new methodologies and intellectual strategies over time;

---

[2] Closely related to this notion is the promotion of universal access, where digital content is disseminated in ways that suit the needs of diverse populations, including the visually or hearing impaired. Program guidelines will make clear to what extent costs for creating transcriptions, translations, or descriptive captions for digital objects are allowable. See discussion of *Allowable and disallowed costs*, below.

3. The digital environment or system could be more cost-effective over time than if a more-traditional path of separated projects individually funded and supported were adopted, because traditional funding of projects is problematic: most, if not all of these projects, are temporarily sustained by grants and/or annual subscriptions or dues levied on institutions of higher learning.

The potential impact and overall scale of the project will necessarily be limited by financial considerations. CLIR and the Foundation are working together under the assumption that only $4 million in re-granting funds would be available for each year the program will operate. An analysis of recent digitization grants given by federal agencies suggests an average award size of just under $150,000 per project. If the proposals to CLIR are similarly sized, it would be reasonable to expect around twenty-seven awards per year. Since the proposed program will emphasize collaborative, multi-institutional approaches, CLIR expects an average request size that is somewhat higher, possibly $200,000 per award. This higher average would result an estimated twenty awards per year.

Officers at the National Historical Publications and Records Commission (NHPRC), which operated a digitization funding program for several years, report using a range of $1 to $3 per scan in assessing the cost-effectiveness of projects creating digital images. This program did not accommodate the wide range of rare and special formats that CLIR expects to see in its proposals in 2015, however. There will undoubtedly be considerable variation in costs across CLIR's funded projects, depending upon original formats to be digitized, requirements for special handling for those formats, the level of descriptive detail in metadata to be created, approaches to quality control and optimization for access (such as optical character recognition), etc., not to mention the effects of inflation, geographic location, and recent improvements in hardware and software tools used for digitization work. It is unsurprising that due to the complexity of these interrelated factors CLIR's investigations have uncovered very little reliable information about average costs-per-item for digitization. Even average cost figures for digitization grants by format are unreliable, since digitization funded through grants often accompanies related research, cataloging, or outreach activities, making cost comparisons across projects extremely difficult without access to the details of project budgets. Due to privacy restrictions budget details are usually not public information.

At the same time, it is possible to find some examples of projects that can serve as tentative benchmarks for estimates of the potential output of CLIR's program. Staff working on the Aldo Leopold Archive digitization project based at the University of Wisconsin, completed in 2008, documented a cost of $2.50 per digitized page. Since this archive includes both print and manuscript archives as well as photographs, its cost may be more representative of the type of work CLIR's program would fund than the lower end of the NHPRC range. Assuming a $200,000 budget, by this measure it would be reasonable to expect a project to digitize 80,000 pages or photographs, or fewer, depending upon related work necessary to make these files

accessible. Audio or audiovisual digitization costs are naturally much higher, often because most time-based media must be digitized in "real time," although this challenge can be surmounted with equipment built to accommodate "parallel transfer" (simultaneous digitization of multiple items).[3] In an academic setting with access to proper equipment and student labor, costs can be relatively modest. Since 2012, the University of Minnesota Libraries has charged faculty, staff, and students $12.00 to digitize one hour of audio or video; external clients pay $25.00 per hour digitized.[4] It is important to stress that, at best, these estimates of average project output are only tentative. By mid-2015, access to the detailed information provided in project proposal budgets should put CLIR in a much better position to describe the size of the potential output of this program.

As the competition proceeds, its relationship and potential contribution to other large scale projects will be continually investigated, as we seek out collaborative opportunities that foster cooperation and allow the various institutions and organizations that will eventually participate in the digitization program to thrive and be sustained within a system of coherently organized scholarly communication.

Section III. Program scope and requirements

The following section serves as a framework for developing program guidelines and the components of the application. A full draft outline of application components is presented in **Appendix 3**. Draft questions for reviewers of applications are listed in **Appendix 4**.

III. i. Application components

The WizeHive online application system currently in use for *Cataloging Hidden Special Collections and Archives* will be adapted for the new program, with significant changes made to the application itself. Applications to the new program will consist of the following components:

Applicant Information. This will include the contact name and details for the individual responsible for the application and, if different, the proposed project Principal Investigator(s); the name and address of the lead institution; any necessary documentation of that institution's eligibility for funding; and the names of project partners.

Project Summary. This will be a brief description of the proposal, to be used for reference in review and in the public documentation of the program on CLIR's website.

Description of Nominated Materials. This will be a more detailed description of the quantity and extent of materials within the collection or collections to be digitized, including an accurate

---

[3] See the Indiana University case study described in Ricky Erway's 2011 report for OCLC Research, *Rapid Capture: Faster Throughput in Digitization of Special Collections*.
[4] See the University of Minnesota Libraries digitization prices here: https://www.lib.umn.edu/digital/pricing

account of their provenance, arrangement, level of available description, current condition, relevant restrictions on access, and intellectual property status. Because of the program's emphasis on comprehensive digitization, applicants will be required to nominate whole collections or substantial proportions of collections to be digitized in their entirety. The descriptions provided in this section must be sufficiently detailed to justify this comprehensive treatment, and they must include estimates of the numbers of items of each format contained in the collection(s), as well as the basis upon which these estimates were determined. CLIR will implement a formula in the online application system that will combine these estimates with the overall budget figures provided in each application so that reviewers will be able to compare rough costs-requested-per-item across proposals.[5] Applicants will describe their approach to the conservation and preservation of the original artifacts and will be required to confirm they understand they hold responsibility for any costs related to that conservation and preservation arising during the proposed project. They will also be required to confirm that they are willing to assume responsibility for the preservation of the digital files created through the project as well as any risks related to copyright infringement, with the understanding that all funded institutions will be required to comply with the program's intellectual property policy and to sign a binding agreement asserting that neither CLIR nor the Foundation holds legal responsibility for project activities [See **Section V. Intellectual property**]. Should the nominated materials still be protected by copyright, applicants would be required to state their rationale for justifying project activities as exceptions under Sections 107 or 108 of U.S. Code Title 17.

Value and Significance. This will be a narrative describing the potential impact of the project upon scholarship. Letters of support from scholars familiar with the collections will be required to attest to the ways the digitized materials will be used for research and teaching. The narrative should also make clear to what extent the materials are currently "hidden" to potential users and include a justification for why digitization (as opposed to cataloging only) is the most appropriate means to maximize the value and significance of the content for scholarship.

Project Context and Impact. Applicants will describe all envisioned deliverables, how each will be made available to the public, and their strategy for outreach to scholars and professionals so that these outcomes have maximum impact both during and after the project. They will identify any preceding initiatives that have informed their approach; they will describe any ways the approach is innovative and how it is cost-efficient; and they will describe any related future initiatives made possible once work is complete. They must demonstrate an awareness of any related collections held by institutions not participating in the project. Applicants will confirm

---

[5] Note that costs-requested-per-item is distinct from actual costs-per item. Accurate estimates of costs-per-item would necessarily include consistent calculations of overhead and institutional investment as well as amounts requested. It would be impossible to achieve this consistency across project proposals without imposing overly burdensome documentation tasks upon applicants. An economic analysis of data reported by grant recipients will ultimately be more helpful in assessing program output and its true costs. Additional factors would be required to properly assess the program's overall impact, including usage statistics, scholarly outcomes, new institutional practices and procedures, etc.

that they are willing to dedicate project metadata to the public domain and to make digitized content available to the public for free, with no fees or new restrictions upon use imposed by the holding institution.

Project Design. This will include a detailed project plan with timeline, the identities and qualifications of key staff, and a description and justification for the proposed workflow that clearly identifies all tools, systems, standards, and technologies to be employed. Applicants should clearly explain the roles and responsibilities of all project partners, as well as describe any training necessary for staff to undertake the proposed work.

Sustainability. Applicants will document and justify their approach to digital preservation, their plans for making digitized content accessible, and their plans for maximizing discoverability through sharing this content and its metadata through appropriate aggregation services. They will be asked to confirm that they understand the program's requirement that all metadata created through funded projects be dedicated to the public domain and that no fees or restrictions upon use be imposed by the recipient institution beyond those already required by law and/or donor agreements. Applicants should also explain how they will disseminate information about any known intellectual property rights and restrictions related to the content so that this information is easily discoverable for users.

Institutional Capacity. Applicants will describe their level of preparedness to undertake project work and document the support of institutional administrators for the proposal. For collaborative initiatives, including any projects involving commercial service providers, applicants will also describe their rationale for partnering for the project, noting any pre-existing relationships with partners or vendors and describing general past experience with inter-institutional collaboration.

Funding. Applicants will submit both a budget and budget narrative that includes cost calculations and justifications of listed costs. Itemized estimates from proposed commercial partners or subcontractors must be appended to budgets, when applicable. Applicants will be asked to submit budgets reflecting only those funds being requested; this represents a departure from previous practice in the Hidden Collections program where applicants were asked to show both funds requested and any institutional cost-share contributions in their budgets. Although cost share figures in a budget can prove useful for reviewers as they evaluate an institution's commitment to a project, there are also disadvantages to this approach. Internal policies dictating what may be proposed as cost share in a formal grant application differ widely from institution to institution; CLIR has found that some applicants are not permitted to include cost sharing in grant budgets regardless of non-grant-funded contributions, while others include indirect cost rates that vary greatly. Additionally, when those awarded grants later submit financial reports, many institutions include cost sharing figures in their project budgets as part of their financial statement, thus complicating efforts to track grant expenditures. Cost share contributions have never been a required element of a request for funding for CLIR's program, so their removal

from the budget does not materially affect the application process. Applicants will still be encouraged to note any institutional contributions in the text of their application, notably within the budget narrative.

III. ii. Allowable and disallowed costs
CLIR has learned through the cataloging program that there is truly no "one size fits all" approach to creating access to collections. This remains true for projects that incorporate digitization. Among CLIR's recipient institutions who have undertaken digitization of their collections (funded through other means), several have chosen to digitize within the same workflow or even digitize before cataloging begins. As examples, Northeast Historic Film (2009, 2010) and the WGBH Educational Foundation (2010) are two recipients that incorporated digitization into their cataloging workflows for their film and video collections, while the University of Michigan and Northwestern University (both funded in 2008) chose to digitize manuscripts and posters first to facilitate more efficient description. Such projects have been among the most innovative and cost-effective approaches to maximizing scholarly access. For that reason, CLIR would like to allow proposals that would include description as well as digitization activities within the same project plan.

A secondary benefit to opening the competition to a variety of workflows would be minimizing the need for CLIR staff to spend time consulting individually with applicants or deliberating with reviewers about whether proposed project collections have been sufficiently cataloged to qualify for the program, or whether particular activities related to further description of those collections are allowable. The majority of finding aids for archival materials do not include item-level descriptions. It is reasonable to assume that digitization projects for even those collections that have already been described in a finding aid or in a catalog at the collection or series level will require the production of some original descriptive metadata, in addition to the technical and administrative metadata required to manage the digital objects created from those materials. The distinctions between the work required to digitize materials that are unprocessed and un-cataloged and that required to digitize materials that have been completely arranged and described are matters of degree rather than precise delineations, and the ability to plan an effective approach to digitization depends upon myriad circumstances, including the nature, current arrangement, and degree of homogeneity of formats in a collection. For these reasons, we feel that constructing a clear definition of a cataloged, "digitization ready" collection that could be applied in all circumstances would be impossible.

Rather than distracting applicants and reviewers with questions about whether particular description activities are allowable or disallowable, CLIR proposes to let applicants make the case that they have sufficient intellectual and physical controls to justify comprehensive digitization of their nominated collections and that they plan to take an approach to their work that maximizes efficiency, discoverability, and access. They must convincingly argue that their collections are "hidden" in the sense that they cannot be exploited for important scholarly work

until they are fully digitized, discoverable and accessible, and they must demonstrate to reviewers that their planned approaches to description and digitization are reasonable and cost-effective, without replicating past work on the collection or wasting time and attention upon unnecessarily detailed metadata.

Allowable costs for the program will principally include labor, services, and training directly related to the mass digitization of collections. Costs associated with creating and sharing original descriptive and administrative metadata for the electronic files created in the course of projects will be considered to be directly related. In cases where collections are unprocessed and/or have limited or no pre-existing catalog records or finding aids, applicants will be allowed to request a limited amount of support for labor, services and training directly related to arranging and describing those collections so long as they can demonstrate that digitization is clearly warranted and these arrangement and description activities can be most efficiently undertaken within the context of the overall digitization project. Costs associated with the retrospective conversion of analog records into digital form will be also be allowable in cases where applicants demonstrate that conversion will be most efficiently accomplished during the proposed project workflow and that such costs are strongly justified and deemed reasonable by reviewers. However, since the emphasis of the program will be on digitization and facilitating access, CLIR will caution applicants that, all else being equal, reviewers will prioritize projects that maximize the investment of grant funds in these areas. Applicants who are in the position to fund cataloging, processing, and retrospective record conversion activities themselves will be strongly encouraged to do so.

Costs associated with making digital content accessible through subject, regional, national, and international aggregation services and repositories will also be allowable, as may be reasonable costs associated with making the content universally accessible. Costs associated with search engine optimization to enhance discoverability will be allowable, as well as storage and backup services performed during the term of the grant. In both cases, applicants will be required to demonstrate the capacity to assume these costs themselves at the conclusion of the project. In the interest of promoting healthy strategic partnerships, costs for travel and shipping directly related to project work should also be allowable, when justified and deemed reasonable by reviewers.

In keeping with the program's emphasis on sustainable and collaborative approaches, purchases of computer software, hardware, or other technical equipment will be limited to a maximum of $5,000 per project. Applicants will be encouraged to share and reuse existing equipment and resources wherever possible. No costs associated with the conservation and preservation of original materials will be funded through the program, including any materials needed for rehousing, but when necessary to undertake project work applicants may choose to list such costs in budget narratives as a demonstration of institutional investment in the project.

As in the *Cataloging Hidden Special Collections and Archives* guidelines, requests for funds to support labor and services provided by an applicant institution's current permanent staff will only be allowable when convincingly justified.

III. iii. Eligibility

For 2015, the *Digitizing Hidden Special Collections and Archives* competition will be open to not-for-profit and educational institutions nominating collections of national significance to scholarship on any subject, in any format. Depending upon the response to the initial request for proposals, CLIR may in future years consider limiting eligibility requirements by subject or format in order to ensure a fair and comprehensive review of applications within the limitations of the program's operational budget and staffing. Any change of this nature would be made in close consultation with the Foundation and the program's review panel.

The Foundation and CLIR aspire to encourage the participation of U.S. based educational and not-for-profit cultural heritage institutions of all types and sizes, and, if possible, international supporting partner organizations to these U.S. institutions. While the participation of institutions outside the United States is highly desirable, due to CLIR's limited capacity to administer grants it is necessary to require that all projects be led by U.S.-based institutions that take responsibility for the direction and fiscal management of each award. CLIR plans to allow only Canadian partners in the first year of the competition and to consider expanding eligibility beyond the U.S. and Canada in subsequent years if it becomes feasible given the program's limited capacity.

Section IV. Application and review activities

CLIR proposes to model the application and review process upon the approach taken for *Cataloging Hidden Special Collections and Archives*. Its principal instrument has been a standing panel of expert reviewers representing diverse constituencies relevant to the program. The panel would ideally be comprised of fifteen members.

Panelists might continue to serve in future years, depending upon their inclination, time, performance, and need for their expertise as determined by CLIR's staff.

As before, the panel will meet twice annually. At the first meeting, reviewers will assess and comment upon all proposals and identify the proportion of the pool that is competitive for funding. At the second meeting reviewers will make recommendations for awards. Deadlines will be set to allow staff two weeks for processing applications and four weeks for reviewers to read proposals and prepare comments. Prior to the first meeting, applicants will submit most of the required documents, exclusive of letters of support, resumes, and detailed technical documentation. Complete proposal packages will be required at the second deadline from applicants approved by the panel to proceed in the competition. More detail about the requirements for each stage of the competition is provided in **Appendix 2**.

CLIR staff will review all proposals for eligibility and compliance with program guidelines following each deadline. Any proposals deemed ineligible or non-compliant will not be sent to reviewers. Ideally, all proposals will be read by at least two reviewers in the first round of competition, and three reviewers in the final round. Due to the open nature of the competition, it is possible CLIR will experience difficulties adhering to these standards using a panel of the size and composition described above. Should an unusually large number of applicants respond to the request for proposals, CLIR will need to implement an internal screening process to remove the weakest proposals from consideration. This may require adjustments to the calendar given below.

Section V. Intellectual property

Since the principal outcome of the proposed digitization program will be digital copies of content that could potentially be protected by law, CLIR will adopt and prominently display a model agreement clearly outlining responsibilities with respect to intellectual property that all grant recipients will be required to sign in order to receive awards. These agreements will be between the recipient institutions and CLIR, and enforceable by CLIR.

CLIR will sign an additional intellectual property agreement with the Foundation. This agreement will state that CLIR is responsible for administering and enforcing intellectual property agreements with recipients of the program's awards. It will require that:

(1) CLIR advise all applicants that payment of grant funds shall be conditioned on the execution of the program's Intellectual Property Agreement with any recipient;
(2) CLIR require receipt of an executed Intellectual Property Agreement from any recipient before granting an award;
(3) CLIR enforce the terms of all program-related Intellectual Property Agreements and bear the costs of such enforcement; and
(4) CLIR seek the written approval of the Foundation before designating another nonprofit entity to receive the license rights granted to CLIR under each Intellectual Property Agreement.

CLIR may request that the Foundation step in and assist with enforcement in particularly difficult cases. The Foundation Office of General Counsel and Secretary will provide assistance and training to CLIR staff in order that CLIR that can administer and enforce the Intellectual Property Agreements rigorously and efficiently.

Appendix 1.
Budget Figures and Justification

[REDACTED]

Digitizing Hidden Special Collections and Archives   *p. 13*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

Appendix 2.

Planning Proposal Narrative, Agendas, and Bibliography related to the ongoing research and consultation process to inform program design

2.1 Planning Proposal Narrative

Research and planning for a new program to digitize hidden special collections and archives
Submitted to The Andrew W. Mellon Foundation
by the Council on Library and Information Resources
January 29, 2014

Contents

1. Summary

The Council on Library and Information Resources (CLIR) proposes to engage in a ten-month research and consultation process to develop the scope and draft and implement guidelines for a new grant program, the successor to the Foundation-funded Cataloging Hidden Special Collections and Archives initiative. For the proposed activities, two types of support are requested. First, CLIR requests $49,702 for additional program staff to assist with administrative duties, freeing program officers' time to engage in research. Second, CLIR requests that the Foundation cover travel and any catering costs associated with the two planning sessions in Washington, D.C. described in Section 3.2, below. The Foundation will fund these meetings directly in accordance with Foundation policies and procedures, so CLIR will not require a further grant to complete this project.

2. Background

The inaccessibility of large quantities of unique materials in cultural heritage institutions' backlogs is a major impediment to scholarship. Begun in 2008, CLIR's Cataloging Hidden Special Collections and Archives program has addressed this problem by supporting and promulgating efficient, innovative approaches to the description of scholarly resources. Now entering its seventh year, the program has funded 109 projects, supporting the cataloging and processing of a rich variety of materials within a wide range of institutions and making them

Digitizing Hidden Special Collections and Archives   p. 14   Program Proposal
Submitted September 17, 2014   Appendix 2. Planning Proposal Narrative, Agendas,
and Bibliography related to the ongoing research
and consultation process to inform program design

newly accessible to scholars, students, and the general public. On December 18, 2013, the Foundation granted CLIR operational funds to support a final round of cataloging projects; these awards will be made in 2014. Projects selected for the program in this final cycle will continue through August 2018.

The original vision for Cataloging Hidden Special Collections and Archives was to promote broad access to special collections and archives within an orderly, flexible, and deeply integrated digital research environment. At the time it was created, CLIR and the Foundation made the commitment to establishing innovative and efficient methods for describing collections because they strongly believed this was a necessary first step toward the realization of this vision. But much more work remains to be done. Reaching the true potential that rare and unique materials hold for scholars will require that their full content be as accessible as possible within current technical and legal limitations.

Fully incorporating special collections and archives into the digital research environment demands the development of new methods and tools for creating and accessing digital surrogates. These methods and tools must complement best practices for cataloging and description. They must be adaptable to a wide range of material types and an equally wide range of institutional contexts. They must be commensurate with the long-term preservation of physical originals as well as their digital derivatives, and they must be sustainable within the regular operating budgets of cultural heritage organizations. They must ensure the exposure of digital objects within a rich yet easily navigable online environment that is responsive to individual researchers' needs. Workflows for digitization must be highly efficient while adhering to standards that facilitate data sharing and interoperability among digital repository systems. Meeting these difficult challenges will require new innovations in practice and open, ongoing communication among curators, librarians, and archivists.

Promoting innovation and efficiency within communities of practice for the benefit of scholarship is both in keeping with CLIR's vision statement, "to transform the information landscape to support the advancement of knowledge," as well as fundamental to the priorities emphasized in the program's founding document. CLIR anticipates that expanding the range of opportunities available to constituents of the Hidden Collections program will significantly increase its impact: while funded projects will continue to make "hidden collections" newly discoverable, these materials will also be made available from homes, offices, or classrooms. Their immediate accessibility will accelerate the pace of scholarship and the deep integration of original sources into teaching and learning. Collections made available online will not just serve scholars' convenience but also open up new avenues of investigation through the application of computational toolsets that support text mining, image analytics, or other methods of extracting data from digitized collections.

Digitizing Hidden Special Collections and Archives   *p. 15*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

Expanding the Hidden Collections guidelines will not be a straightforward exercise. A re-examination of the program's core values in the light of current best practice in the digitization of special collections and archives will be required. While some aspects of the program, such as its eligibility requirements, could remain unchanged, it will be necessary to adjust the range of allowable costs, to set new benchmarks for the examination of project plans, and to compose new application questions and perhaps new review procedures.

While CLIR officers have significant experience with the review and administration of grants to cultural heritage institutions, they have limited familiarity with digitization workflows appropriate for unique and rare materials. For that reason, CLIR proposes to conduct a review of current literature and active practice in the digitization of special collections and archives while at the same time engaging in a two-step consultation process in cooperation with the Foundation.

3. Project Description

CLIR proposes to conduct the first phase of the proposed research in two parallel workstreams. The first of these will be an environmental scan that will collect information relevant to the digitization of cultural materials. This scan is described below in Section 3.1, illustrated by an example bibliography. The second will entail planning and holding two consultation meetings with outside experts in Washington, D.C. The goals and outcomes for these meetings are below in Section 3.2. Both the environmental scan and consultation meetings will occur in the spring and early summer of 2014.

During the second phase of the project, CLIR will translate phase one findings into draft guidelines that will become the foundation for a new program for describing and digitizing hidden collections.

3.1. Environmental scan: CLIR program officers Christa Williford and Jena Winberry will conduct the environmental scan with the support of Amy Lucko. Their research will begin immediately upon approval of this proposal and will continue through July, when CLIR will submit to the Foundation its prospectus for operational funds for the new program. The scan will collect together information on the following topics:

1. Current national-level funding programs that support the digitization of cultural materials;
2. Documentation of current best practices for digitization of cultural materials, focused principally on digitized manuscripts, printed materials, photographs, and other fixed media;
3. Documentation of current resources, tools, and best practices for the preservation of digital text and image formats;
4. Documentation of common metadata standards used for digitized cultural materials;

Digitizing Hidden Special Collections and Archives   *p. 16*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

5. Current metrics that describe the time and costs required for digitization of cultural materials according to standard best practices;
6. Any notable accounts of differences between library/archival and museum practices, standards, or cost models related to digitization;
7. Any notable accounts of differences between practices, standards, or costs for digitization by external vendors and those for digitization conducted on-site.

The following are examples of resources that may inform the extension of the Hidden Collections program; they serve as a beginning point for the environmental scan.

3.1.1. Grant programs
- *Humanities Collections and Reference Resources*. Division of Preservation and Access, National Endowment for the Humanities (NEH). Most recent request for proposals at: http://www.neh.gov/grants/preservation/humanities-collections-and-reference-resources
- *Digitizing Historical Records*. National Historical Publications and Records Commission (NHPRC). Most recent request for proposals at: http://www.archives.gov/nhprc/announcement/digitizing.html

3.1.2. Reports
- Erway, R. (2011). *Rapid Capture: Faster Throughput in Digitization of Special Collections.* OCLC Research. Retrieved from http://www.oclc.org/content/dam/research/publications/library/2011/2011-04.pdf.
- Lee, C. A.; Woods, K.; Kirschenbaum, M.; and Chassanoff, A. (September 2013). *From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions*. A product of the BitCurator Project. Retrieved from http://sils.unc.edu/news/2013/bitcurator-white-paper
- Maron, Nancy L. and Pickle, Sarah (November 2013). *Searching for Sustainability: Strategies from Eight Digitized Special Collections*. Ithaka S+R and the Association of Research Libraries. Retrieved from http://www.arl.org/publications-resources
- McGovern, N. and Skinner, K., eds. (2012). "Aligning National Approaches to Digital Preservation." Educopia Institute Publications. Retrieved from http://www.educopia.org/sites/educopia.org/files/ANADP_Educopia_2012.pdf.

3.1.3. Conferences and events
- Museums & the Web 2014 (Baltimore, MD): http://mw2014.museumsandtheweb.com/program/
- iPres 2014 (Melbourne, Australia): http://ipres2014.org/
- JCDL (London, UK): http://www.dl2014.org/
- Digital Humanities 2014 (Lausanne, Switzerland): http://dh2014.org/
- Digital Preservation 2014 - NDIIPP & NDSA (expected: VA)
- 2014 DLF Forum (Atlanta, GA): http://www.diglib.org/forums/

Digitizing Hidden Special Collections and Archives   *p. 17*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

- LODLAM Summit 2014 (TBD): http://lodlam.net/
- CNI Spring 2014 Meeting (St. Louis, MO): http://www.cni.org/event/cni-spring-2014-membership-meeting/

3.1.4. Relevant articles and other resources
- Bailey, Charles W., Jr. (2010-2013). *Digital Curation and Preservation Bibliography*. Retrieved from http://www.digital-scholarship.org/.
- Bak, G. (2012). "Continuous Classification: Capturing Dynamic Relationships among Information Resources." *Archival Science* 12.3: 287.
- *BitCurator: Tools for Digital Forensics Methods and Workflows in Real-World Collecting Institutions*. http://www.bitcurator.net/publications/.
- *Digital Preservation in a Box*. National Digital Stewardship Alliance. http://dpoutreach.net/
- *Federal Agencies Digitization Guidelines Initiative* (2007-present). Retrieved from http://www.digitizationguidelines.gov.
- Lazorchak, B. (2012). "A Digital Asset Sustainability and Preservation Cost Bibliography." *The Signal: Digital Preservation*. Retrieved from http://blogs.loc.gov/digitalpreservation/2012/06/
- Miller, K. (2013). "All Text Considered: A Perspective on Mass Digitizing and Archival Processing." *The American Archivist* 76.2: 521-541.
- *NDSA National Agenda for Digital Stewardship* (2014). National Digital Stewardship Alliance. Retrieved from http://www.digitalpreservation.gov/ndsa/nationalagenda/.
- Sustainability of Digital Formats: Planning for Library of Congress Collections (2004-2013). Retrieved from http://www.digitalpreservation.gov/formats/index.shtml.
- Sutton, S. (2012). "Balancing Boutique-Level Quality and Large-Scale Production: The Impact of 'More Product Less Process' on Digitization in Archives and Special Collections." *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13.1: 50-63. Retrieved from http://rbm.acrl.org/content/13/1/50.full.pdf+html.
- Zhang, J. and Mauney, D. (2013). "When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation." *American Archivist* 76.1: 174-195.

3.2. Planning sessions: CLIR proposes to hold two half-day planning sessions in Washington, D.C. in the spring and early summer of 2014. The purpose of the first of these meetings (tentatively planned for March) will be to define the scope and priorities for the new program, and to identify ways it will be distinct from and complementary to other national-level funding initiatives. The purpose of the second of these meetings (tentatively planned for May or June) will be to establish key criteria and guidelines that will be directed toward the most urgent needs of program constituents. Ultimately, these criteria and guidelines should be easy to interpret for

Digitizing Hidden Special Collections and Archives   *p. 18*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

applicants and reviewers, so engagement with consulting experts from diverse backgrounds will be critical to success.

3.2.1. Scope and priorities planning session: For this meeting, CLIR will convene leaders of national organizations in the humanities, senior representatives from relevant funding agencies, and directors of large-scale digital humanities initiatives. The outcomes of this meeting will be

1. a statement about the scope and priorities for the new program (the nature of materials to be digitized, the expected minimum requirements for project eligibility and outcomes, etc.);
2. an explanation of how CLIR's program will be distinct from and complementary to the priorities of other funders as they relate to the digitization of cultural heritage materials; and
3. a set of more detailed questions for further exploration that will comprise an agenda for the second planning session to be attended by experienced practitioners.

3.2.2. Criteria and guidelines planning session: For this meeting, CLIR will convene knowledgeable professionals in relevant areas such as mass digitization of special collections and archives, digital preservation, metadata production and extraction, and the management and long-term sustainability of digital libraries and archives. The outcomes of this meeting will be

1. a set of answers to the questions developed during the first planning session that represent current thinking about best practices in description and digitization for special collections and archives;
2. a framework based upon these answers within which CLIR can shape the guidelines for the new program (minimum and maximum award amounts, allowable and disallowed costs, etc.); and
3. a set of review criteria to be used to assess proposals submitted to the new program.

3.3. Staffing

[REDACTED]

3.4. Timeline: The duration of the proposed research and consultation process is ten months, beginning March 1, 2014, and concluding December 31, 2014. Additional activities related to the effort will occur both before and following these dates.

January-February 2014: Environmental scan; pre-planning for project and preparation to hire new administrative assistant. In the event that the start date for the new assistant precedes the start date for the proposed project (March 1), CLIR will cover the assistant's salary and any other associated costs related to the position until that date.

Digitizing Hidden Special Collections and Archives   *p. 19*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

March 1, 2014: Formal project start date

March 2014: Scope and priorities planning session (Washington, D.C.). This meeting will be funded by the Foundation, and no funds are requested by CLIR for this purpose in this proposal.

March-April 2014: Environmental scan continues; planning for second consultation session; Administrative Associate joins CLIR staff to assist with the administration of the current program while program officers continue research activities

Early May 2014: CLIR Hidden Collections Review Panel Meeting (Washington, D.C.).

May or June 2014: Criteria and guidelines planning session (Washington, D.C.). This meeting will be funded by the Foundation, and no funds are requested by CLIR for this purpose in this proposal.

July 15, 2014: Deadline for initial prospectus for the new program

August 15, 2014: Deadline for draft proposal for the new program

September 15, 2014: Deadline for completion of final proposal for the new program

December 31, 2014: Formal project end date

January 2015: Request for proposals for the new program (pending approval by the Foundation)

March 31, 2015: Deadline for submission of final narrative and financial reports on the current planning proposal

4. Outcomes

The results of the environmental scan and consultations will be compiled into a proposal for the creation of the new grant program. An initial prospectus will be submitted to the Foundation on or before July 15, 2014; a first draft on or before August 15, 2014; and a final draft proposal will be submitted no later than September 15, 2014. The proposal will include all elements enumerated in the Foundation's guidelines for its Scholarly Communications and Information Technology (SCIT) Program.

Additional outcomes to be prepared in the latter half of 2014 include a new application form and review portal; new guidelines and instructions for applicants, including a presentation to be used in applicant webinars and for other public speaking engagements related to the program; a new

Digitizing Hidden Special Collections and Archives   *p. 20*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

review panel for the program; guidelines and instructions for these reviewers; and revisions to all content on CLIR's website related to the Hidden Collections initiative.

5. Data Collection and Intellectual Property

The outcomes of the proposed effort will be as freely shared as possible, following the advice of Foundation officers and other participants. As most of the research should involve public information, CLIR does not anticipate collecting significant amounts of sensitive data; however, should any expert share non-public information (for example, applicant demographics or statistics for a particular grant program) that the expert would prefer CLIR keep confidential, CLIR will honor those wishes.

The new application form developed during this process will be designed using the WizeHive service. The application form and guidelines will be made freely available to the public on CLIR's website, in the event the evolution of the Hidden Collections program is approved by the Foundation in late 2014. WizeHive is already licensed by CLIR. CLIR will develop no new software tools or applications in the course of this project.

Intellectual property produced during the course of any future grants made through CLIR's new digitization program will be subject to the Foundation's intellectual property policy.[1] At the criteria and guidelines planning session to be held in Washington, D.C. in May 2014, CLIR will consult with the Foundation's legal counsel regarding intellectual property agreements for the proposed digitization grants.

6. Sustainability

Data gathered and reports produced during the course of research will be maintained by CLIR on its shared server through at least the calendar year 2015, so that it might inform the creation and first year of development of the new program. Materials will be deleted from CLIR's servers once no longer relevant, or in the event the program is discontinued.

2.3 Digitizing Hidden Collections: Select Bibliography, 6/13/14

Atkins, Winston et. al. *Staffing for Effective Digital Preservation*. NDSA, 2013.

Erway, Ricky. *Rapid Capture: Faster Throughput in Digitization of Special Collections*. Dublin, OH: OCLC Research, 2011.

Maron, Nancy L. and Pickle, Sarah. *Searching for Sustainability: Strategies from Eight Digitized Special Collections*. Ithaka S+R and the Association of Research Libraries, November 2013.

Digitizing Hidden Special Collections and Archives   *p. 21*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
                                                         and Bibliography related to the ongoing research
                                                         and consultation process to inform program design

Miller, Larisa K. "All Text Considered: A Perspective on Mass Digitizing and Archival Processing." *The American Archivist* 76.2 (2013): 521-541.

Rieger, Oya Y. "Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives." *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 11.1: 11-22.

Rimkus, Kyle, Thomas Padilla, Tracy Popp, and Greer Martin. "Digital Preservation File Format Policies of ARL Member Libraries: An Analysis." *D-Lib Magazine* 20.3/4 (March/April 2014).

Sutton, Shan C. "Balancing Boutique-Level Quality and Large-Scale Production." *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13.1: 50-63.

Zhang, Jane, and Dayne Mauney. "When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation." *The American Archivist* 76.1 (2013): 174-195.

Other Links:
New draft NHPRC Guidelines for "Online Publishing of Historical Records" Program
Charles W. Bailey's Digital Curation Bibliography and 2012 Supplement
Digital Preservation in a Box
Sustainability of Digital Formats: Planning for Library of Congress Collections
Federal Agencies Digitization Guidelines Initiative
ARL Web Accessibility Toolkit

Digitizing Hidden Special Collections and Archives   *p. 22*   *Program Proposal*
Submitted September 17, 2014   *Appendix 2. Planning Proposal Narrative, Agendas,*
and Bibliography related to the ongoing research
and consultation process to inform program design

Appendix 3.
Digitizing Hidden Special Collections and Archives
Eligibility Criteria and Application Outline

[REDACTED – THIS INFORMATION HAS CHANGED SINCE THE TIME OF ORIGINAL PROPOSAL SUBMISSION. INTERESTED APPLICANTS SHOULD VIEW CLIR'S WEBSITE FOR THE MOST CURRENT INFORMATION.]

Appendix 4.
Draft Questions for Reviewers of Applications

The following questions, once refined, will help program staff prepare advice and supporting documentation for reviewers. The questions will also inform mechanisms for the scoring of applications and for the provision of feedback to applicants. CLIR expects to provide these same questions to applicants for use in self-assessments as they prepare their proposals and to help increase the transparency of the review process.

Scholarship
1. Are you confident the proposed project would have a significant impact on scholarship and teaching?
2. Are you confident that the proposed project would support the creation of new knowledge?
3. Will the proposed project make available materials that are not otherwise available to scholars?
4. Will the availability of project content in digital form make possible new kinds of research, and/or make it possible to ask new kinds of research questions?
5. Will the proposed project make content available, discoverable, and easily usable for the scholarly communities and the general public?

Comprehensiveness
1. Will the proposed project result in the availability of a sufficient quantity of related materials to support research?
2. Do the applicants justify the significance of the entirety of the collections they are nominating for digitization?
3. Do the materials to be digitized relate to a sufficiently broad range of topics of high interest to researchers in multiple disciplines?
4. Do the applicants have a feasible project plan that maximizes efficiency of throughput while providing for appropriate levels of description to support discoverability?

Collaboration
1. Do the applicants have a good command of the broader scholarly and professional contexts for their work, and do they understand the relationships between the nominated collections and other collections at other institutions?
2. Do the choices of project staff and institutional partners (if any) seem appropriate to the ambitions of the project as well as strategic in their service of long-term institutional goals?
3. Does the project plan avoid unnecessary duplication of capacity and effort?

4. Do all the project participants seem prepared to undertake project work?
5. Do the applicants have outreach and marketing strategies for this project that are likely to be effective?

Sustainability
1. Do the applicants understand and are they prepared to adopt appropriate standards, technologies, and practices for ensuring the long-term availability and discoverability of digital files?
2. Are the applicants' plans for digital preservation appropriate for the formats of the materials to be digitized?
3. Does the approach to project work seem especially innovative or cost-effective in ways that could make it a model for others to follow?

Openness
1. Do the applicants seem to understand the legal and ethical issues related to digitizing their nominated materials and are they prepared to address these issues and accept any associated risks?
2. Will the digital files created through the project be made fully available to users, for free? If not, are the strategies for opening access appropriate to the nature of the collections, without any new or unnecessary restrictions, fees, or other barriers being imposed by the applicant institution?
3. Will the applicants display information about intellectual property rights and any attribution requirements related the digital content in a way that is both easily discoverable and comprehensible for users?
4. Have the applicants considered strategies for opening access to their collections to users who are vision or hearing impaired?
5. Are the applicants prepared to dedicate metadata created through the project to the public domain, and do they seem willing to share this metadata in ways that support the aggregation of their content with related collections at other institutions?