



## The Challenges of Sustaining a Long-Term Collaboration: Reflections on the Philadelphia Hidden Collections Projects

---

*David McKnight, Director, The Rare Book and Manuscript Library, University of Pennsylvania*

*Eric Pumroy, Associate Chief Information Officer and Seymour Adelman Director of Special Collections, Bryn Mawr College*

### Abstract

During the past five years, the Philadelphia Area Consortium of Special Collections Libraries (PACSCL) received two CLIR Hidden Collections grants that tested the capacity of this voluntary collaboration of academic and cultural institutions to develop and sustain a unified methodology and provide a single point of public access to their manuscript and archival collections. While the projects have been successful in achieving their goals, there continue to be unresolved issues around the development and maintenance of the project database. This two-part paper looks at these projects as a means of examining the challenges of building a sustainable multi-institution technology-based program, and does so by looking at the issues from the perspective of both the host institution and the participants.

The Philadelphia Area Consortium of Special Collections Libraries (PACSCL) received two Hidden Collections grants to process archival and manuscript collections in the region: one in 2008 for \$500,000, and a second, smaller grant in 2011 for \$249,000. These are the most recent of a long series of collaborative grant-funded projects that this consortium of 36 academic institutions, historical organizations, and museums has received since its founding in the 1980s, including grants from the Pew Charitable Trusts in the early 1990s to address uncataloged book and manuscript collections; a grant from the William Penn Foundation in the late 1990s to set up online catalogs for the smaller institutions in the consortium; and a grant from The Andrew W. Mellon Foundation in 2005 to survey unprocessed manuscript collections and identify priorities for further attention. This last project, which ended in 2008, served to identify most of the collections that were processed in the two CLIR grants.

The two Hidden Collections grants were highly successful in meeting their goals of reducing the backlogs of important unprocessed collections in the region, and in fact, they processed more than had originally been called for in our proposals. In the first grant project, 133 collections totaling 4,000 linear feet were processed across 21 institutions. In the second, smaller project, 45 collections totaling 1,685 linear feet were processed across 16 institutions, some of which were not part of the first project. In all, 178 collections totaling 5,865 linear feet were processed during the two CLIR Hidden Collections projects and involved collections from 24 institutions.

As important as the processing was, it was only a piece of a larger and more ambitious plan to create a finding aids database that would provide access to all historical collections in the region. The University of Pennsylvania Library was in the process of setting up such a database for its own finding aids at the time we were developing



the first proposal and generously offered to expand it to include finding aids from PACSCL institutions. The proposal for the initial grant explicitly stated that the Hidden Collections project finding aids would only be the beginning point for the database, which we saw eventually becoming the critical site for research in the Philadelphia area. Seven years and two projects later, the database has become much richer, but the number of institutions that routinely add new finding aids to it is not as large as we had hoped. The purpose of this paper is to look at this project as a test case of what it means to collaborate on a project that involves building and maintaining a technology infrastructure. The first part will consider how the database has been used and the factors that have determined institutions' level of involvement, and the second will examine the implications of a collaborative technology project for the host institution.

### ► **Part I: The Participants' Perspective**

**Eric Pumroy**, Bryn Mawr College

In many respects, the finding aids database has been an impressive success. As of spring 2015 there are about 2,800 finding aids in it, compared with the 178 finding aids created during the two Hidden Collections projects. Our great surprise is that the largest contributor is a program that did not exist when we got started: The Hidden Collections Initiative for Pennsylvania Small Archival Repositories, funded by grants from The Andrew W. Mellon Foundation and developed and managed through the Historical Society of Pennsylvania. As of March 2015, there were more than 900 finding aids from about 150 repositories represented in the finding aids database, and nearly all of these were collections that were invisible to the larger research community until the project and the database provided a way to catalog and publicize them. The other major

contributors are the University of Pennsylvania with more than 700 finding aids, the Hagley Museum and Library with 671 finding aids loaded in the fall of 2014, and the rest of the PACSCL institutions with about 500. It is this last group that I want to look at more closely. Aside from the University of Pennsylvania and Hagley Museum and Library, the PACSCL members that contributed the greatest number of finding aids were Drexel University, Haverford College, Bryn Mawr College, the Free Library of Philadelphia, and the College of Physicians of Philadelphia. Contrary to our expectations, most others have not continued submitting finding aids. In a survey conducted at the end of the first project in 2012, 17 institutions, or about three-quarters of the participants, said that they were very pleased with the database and would continue to submit finding aids. In late 2014 and early 2015 we did a follow-up survey of all PACSCL members, and of the 17 respondents who had participated in the CLIR grants, 8 said that they planned to continue reporting (for the most part, they are the ones that have been reporting), and 8 said that they were uncertain.

The reasons for the uncertainty and lack of continued submissions vary by institution, but most of them come down to issues with technology. The way the system works is that an institution creates a finding aid using Archivists' Toolkit, and then uploads the finding aid to its own web server where Penn's automated harvester finds it and loads it into the finding aids database. Several of the smaller institutions reported a lack of IT support, which makes running Archivists' Toolkit and uploading finding aids difficult. Others had to deal with institutional IT policies that blocked access to the harvester. There are also problems with the database itself that held institutions back from



loading finding aids, particularly in the way it requires collection numbers to be formatted. And finally, a few institutions' finding aids were dropped from the database when their websites moved and the harvester could no longer locate their finding aids. None of these problems are insurmountable, but solving them requires time, attention, and in some cases, money.

The finding aids database was always intended to be for all collections in the region, not just for those institutions that participated in the Hidden Collections projects. The recent survey was sent to all PACSCL members as a way of capturing how, or if, the non-Hidden Collections institutions thought about the database. These institutions did not sit out the two PACSCL projects because of opposition or lack of interest, but rather because the projects had been based on the results of the earlier consortial survey that ranked the research values of unprocessed collections in the region. As a consequence, the institutions that processed their most important collections on their own did not have collections that qualified. A number of major institutions, such as the American Philosophical Society and Swarthmore College's Friends Historical Collections fell into this category. Because these institutions were not involved in the Hidden Collections grants, they did not give much consideration to the finding aids database, and in fact, several had not realized that contributing their findings aids was an option. Several reported that they would be interested, but would need advice on how to go about it. While these are only a handful of institutions, the responses indicate a continuing interest in using the database as a way of improving access to collections.

One of these institutions, the Hagley Museum and Library, on its own initiative contacted Penn about loading its finding aids on to the PACSCL

site, and its finding aids now constitute nearly a quarter of the finding aids. The Hagley's decision came about as an offshoot of a larger project to create its own EAD finding aids database using the open source software XTF, developed by the California Digital Library. Once the finding aids were in their own database, it was a relatively straightforward matter to also make them available to the PACSCL site. At this point the benefits of having their finding aids in the PACSCL site aren't clear because they have only been there for a few months, but as Library Director Erik Rau reported, they thought that it was important for their finding aids to be findable as part of a larger regional collection and exposed to a wider population of users.

While there is interest among most PACSCL institutions in having the finding aids database succeed, the database is nonetheless a low priority for them. Most institutions have their own ability to mount finding aids so that they are findable through web searches, so adapting their practices to make their finding aids conform to the database standard and putting energy into addressing technical barriers are not seen to be good uses of limited staff and resources. It is hard to argue with this point of view because it is not clear what difference the database has made to the institutions whose finding aids are there, or indeed, if many potential researchers are using it. In part this may be an issue of critical mass; the richer the database becomes, the more essential it will be for researchers working in the Philadelphia region and consequently the more importance institutions will attach to participating. It also has to be actively promoted, both to people doing historical research and within PACSCL.

Which leads us to institutional structure. How does a voluntary organization develop, maintain, and promote an ongoing project that depends



upon a robust technical infrastructure? For this program, the answer has been that we rely on one of our members, the University of Pennsylvania Libraries, to handle nearly all of the work and all of the costs. Penn provides and maintains the hardware, manages the software and the design of the site, and assigns a regular staff member in Special Collections to troubleshoot problems and advise PACSCL members. The staff person, Holly Mengel, was also the director of the first PACSCL CLIR grant, so she knows the system well and is deeply committed to its success. Nonetheless, Penn, not PACSCL, is paying all of her salary. Neither PACSCL nor individual PACSCL institutions provide financial support to help Penn maintain the site, and there is not a formal PACSCL advisory group to help with policies, troubleshooting, and promotion for the site, although there are individuals within PACSCL on whom Holly has been able to call for assistance.

PACSCL has been remarkably successful over the last 30 years at raising money for projects with limited duration, and these projects have had an enormously positive impact on the condition of historical collections in the region. But it is no longer enough just to catalog and process collections, something at which we have excelled over the years. If the cataloging, processing and, now, digitization work is going to be useful for our publics, we also need to figure out how access is going to happen, and that means some form of technical infrastructure that draws the collections together and enables scholars and students to make connections across our holdings. Now, the institutions with enough resources are building those structures individually, and because of the needs of their institutions and users, this is entirely understandable. The problem remains that many of PACSCL's members are independent cultural

organizations that struggle with managing an IT infrastructure that is both increasingly complex and increasingly essential for making their collections visible to the larger world. In the long run, if not immediately, making alliances with better-resourced institutions is becoming essential. For the non-PACSCL members—the 150 historical organizations whose collections are in the finding aids database because of the Historical Society of Pennsylvania project—their collections will be visible only if they can partner with bigger players. Even the wealthier institutions are not always able to draw on IT support in the way they desire. Institutions with large IT departments also have large IT needs that can absorb all of the available energy, leaving limited IT support for special collections work. Finding ways to leverage developments at any one institution so that many more can benefit, in the way that we have done with the finding aids database at Penn, looks like a sensible way to operate.

The challenge will be to find a sustainable mechanism for building a technical infrastructure in an environment of mostly private independent institutions, and without a publicly funded institution, like the California Digital Library, to provide continuity and leadership. Depending on the public service ethic of a few of the larger institutions, especially the University of Pennsylvania Library, has worked reasonably well, but it leaves all of the participants dependent upon Penn's decisions and priorities for the maintenance of a system that many of us are coming to depend upon. Finding ways of providing financial support to Penn seems only fair, but also risks putting Penn in the uncomfortable position of being a vendor. The more money involved, and the more reliance other institutions place in the



system, the more likely that our very informal, low-cost arrangements will need to be replaced by memoranda of understanding and service level agreements. These might be important improvements for the long-term health of PACSCL, but they will also be a major change from the handshake agreements that we have operated with in the past. It also only seems fair and wise for PACSCL institutions to find a way of taking a more active role in overseeing, troubleshooting, and promoting the finding aids database. The support structures that were put in place for the grant projects have now gone away, so reconstituting them in a sensible, sustainable way seems to be the next order of business.

Finally, and not to be forgotten, are our users. Philadelphia is an extraordinary cultural center, with rich collections spread among some of the country's oldest and most distinguished institutions. This wealth of cultural institutions also means that collections are dispersed across the region, making the task of locating all of the relevant sources on a topic a daunting one for researchers. The Philadelphia region is full of situations where family papers are split among several institutions, or records of organizations that deal with similar social, cultural, and political issues are housed in different repositories, or documentation of neighborhood and community life is found in numerous and often unexpected locations. It is true that much of this documentation could be found through web searches, but not as efficiently and not with the detail and nuance that a regional database can supply, particularly for researchers interested in social issues where proper name searches are less productive. For our local users, including the numerous college and increasingly engaged high school students, a regional database has the potential to encourage more ambitious

and comprehensive research by revealing avenues to sources that would otherwise have remained obscure. The CLIR grants have given Philadelphia institutions the opportunity to create an exceptionally useful tool for scholars, students, and the public doing historical research in the area. Our challenge now is to find a way of erecting an organizational and financial structure to keep it going.

## ► **Part II: The Institutional Host's Perspective**

**David McKnight**, University of Pennsylvania

In the previous section, Eric has amply described PACSCL as a “coalition of the willing” and the issues related to managing a large collaborative project of participatory members of the consortium. What I would like to discuss in the second half of this paper are some of the issues related to the pros and cons of administering such a project from the perspective of the host institution—in this instance my own institution: The University of Pennsylvania Libraries.

I would like to preface my remarks by citing another example: [The Digital Scriptorium](#). Unlike PACSCL, the Digital Scriptorium is a national consortium of institutional libraries that have chosen to contribute selected digital images from each institution's collection of Medieval manuscripts. There are currently 36 members of the consortium. What differentiates the Digital Scriptorium from PACSCL is that it does have an institutional host: The University of California Berkeley Library. Without going into great detail, the Digital Scriptorium (DS, hereafter) began its existence at Berkeley in the late 1990s; however, when the DS's executive director, Consuelo Dutschke, relocated to Columbia University the project moved with her including the digital assets and the DS database. In 2010, Columbia University Libraries made it clear that it was no



longer willing to host the DS. With the news that Columbia was abandoning the DS, there was shock and disappointment, especially among the members of the Executive Committee, of which I was a member. Although Columbia was bidding farewell to the housing and maintenance of this noble primary source digital collection, there may have been a sigh of institutional relief. But the library was willing to continue to underwrite Consuelo's salary on a part-time basis so that she could devote her time to her directorship of the DS.

After a two-year search, the DS Executive Committee reviewed several offers to host the DS project, its database, and digital images. The winner was UC-Berkeley. Thus the project that was launched at Berkeley with NEH funding ten years earlier was now returning to Berkeley. The primary challenge of the "new" technology host was to migrate the cumulative data that was housed at Columbia and "cross-walk" the data into a new framework: the California Digital Library. Thus the technology host was located on the west coast and the curatorial center of the project remains on the east coast. Like other long-term digital humanities projects, the DS is faced with ongoing sustainability and maintenance issues: uploading new data, web development, database issues, and data storage, among others. As technology host, Berkeley, like other institutions, has had to off-load the cost of maintaining the DS site by charging membership dues with the goal of raising \$25,000 to pay for the annual maintenance fee. Based on a set of standard metrics, each member of the DS pays a prorated annual fee toward the upkeep of the DS database.

Although the DS has a technology host, the DS has found it difficult to establish itself as a 501(c)3 organization. This makes it nearly impossible for the DS to apply to accept donations or apply for

grants. Thus the DS is in a state of limbo with respect to applying for external funding. (To compound the matter, grants would be administered by UC-Berkeley and subject to California law, which might pose problems for both parties.) PACSCL's situation is not as complex as the Digital Scriptorium's, and it has the advantage of having 501(c)3 status. Nonetheless, PACSCL is still dependent on member institutions to serve as the administrators of major grant projects because it is such a small organization that a large grant would complicate its legal status and overwhelm its ability to manage the project. So if you look at the awards for the 2008 and 2012 PACSCL CLIR grants, the awardee is not PACSCL but rather the University of Pennsylvania. Penn Libraries was a founding member of the consortium and is, if I may say, modestly, among the larger and better-resourced institutions in PACSCL. With that said, at the time that PACSCL was planning to submit its initial application, I volunteered to inquire if Penn might serve as institutional host for the grant itself and also provide the infrastructure for supporting the project team for the duration of the project. Penn Libraries' Vice-Provost and Director of Libraries H. Carton Rogers endorsed the notion of Penn Libraries serving in both capacities to support the first CLIR grant in 2008 and the second grant in 2012. What does institutional support actually mean?

First it meant that Penn Libraries would serve as the 501(c)3 institution that submitted the grant application on behalf of the consortium. During the application authoring process, Penn's role as host institution was clearly defined according the following criteria:

- Project oversight
- Administrative support
- Human resources



- Infrastructure
- Technology
- Long-term sustainability

The project managers for both grants were hired as University of Pennsylvania employees (as were the graduate student processors), but were also expected to report regularly to the PACSCL board. In terms of project oversight, in the case of both CLIR grants there were two principal investigators, one representing PACSCL and the other the University of Pennsylvania Libraries. PACSCL was very fortunate in hiring two exemplary project managers. There was no need to remind the managers that the reputations of both PACSCL and Penn were dependent upon the successful completion of their respective projects.

Institutional support meant more than just providing office space, lights, heat, an Internet connection, a photocopier, and office supplies. These were included in the cost-sharing portion of the project budget. What was not included was the time of human resources staff, who assisted with the posting of jobs and payroll issues, and the library's business office, which managed the grant funds and generated the financial reports for inclusion in our interim and final reports.

Finding offices in a building with few free office spaces for staff proved challenging but not insurmountable; the one area in which Penn played a leading and essential role was technology. Apart from buying project laptops and setting up or moving phones and computers (during CLIR I the project staff were moved to another location in library because of construction on the sixth floor of Van Pelt Library), technology proved to be the most demanding aspect of these projects.

Penn had implemented its digital library application to search, retrieve, and display XML-encoded data, and through it, the library's technology unit had implemented a Penn EAD repository. Extending the model to the PACSCL project with 20 participants as opposed to Penn's three separate units was perceived by the Library Information Technology group as an opportunity to develop the library's digital library technology staff. It should be noted that in both CLIR budget requests we included a modest five percent of the total grant for technology support. The actual cost in terms of coding labor has not been quantified to date. As Penn finessed the EAD repository, the project managers took care of site technology problems: setting up local instances of Archivists' Toolkit, creating web folders, training staff, and troubleshooting. It was part of their job description.

Although the two CLIR PACSCL Hidden Collections processing grants have been completed, the question of whether Penn will continue to serve as host to the EAD PACSCL finding aids database has not been resolved. The data are secure. The finding aids are searchable and accessible. But what is the cost of long-term maintenance? I refer not to the actual data itself, but the tweaks to code, code errors, improvements, and possible migration of data to a new platform. These represent challenges for both PACSCL and the University of Pennsylvania to resolve over time.

## Conclusion

This is a time of transition for both PACSCL and historical collections management as a whole. For many years our emphasis has been on reducing backlogs of uncataloged and unprocessed collections and placing the catalog records and finding aids into both our



institutions' library systems and the handful of national systems, particularly WorldCat. For the last 25 years, PACSCL has been a model of how a private group of cultural institutions can work together effectively to raise money to address backlogs and professional standards and capabilities in a major metropolitan area. The success in Philadelphia and nationwide in reducing backlogs, coupled with the growing amount of digitization, has now changed the landscape for cultural organizations, for with increasing amounts of collections metadata and digital images come new opportunities for scholars and students to access collections and undertake research in ways that were difficult in the past. This new research will depend upon easy access to data, and preferably access that draws collection data from across institutions in easily searchable databases.

We are already seeing numerous structures being created that take advantage of collections data from multiple institutions. In addition to the PACSCL Finding Aids Database, Penn supports the [Provenance Online Project](#) and the [Early Novels Database](#), and has just brought up the [OPenn site](#) that will host the digital collections in the PACSCL diaries project currently under way, and the proposed collections of medieval

manuscripts housed in the region. Outside of the Philadelphia region, Bryn Mawr has led an initiative among the schools formerly known as the Seven Sisters College to create a [mechanism for searching the digitized collections of student letters, diaries, and scrapbooks](#), and there are many more multi-institution projects under way in other regions. At this time of excitement over digital technologies and their potential for transforming the ways in which scholars and students interact with cultural collections, it is understandable and laudable that libraries and archives are experimenting with developing new structures for providing access to their holdings. Finding a means of sustaining the most important of these structures will be the great challenge of the coming years.

## References

College Women: [www.collegewomen.org](http://www.collegewomen.org)

Digital Scriptorium: <http://bancroft.berkeley.edu/digitalscriptorium/>

Early Novels Database: <http://earlynovels.org/>

OPenn site: <http://openn.library.upenn.edu/>

Provenance Online Project:  
<https://provenanceonlineproject.wordpress.com/>