

**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 uncatalogued book and manuscript collections; a grant from the William Penn Foundation in the late 1990s to set up online catalogues 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 4000 linear feet were processed across 21 institutions. In the second, smaller project, 45 collections totaling 1685 linear feet were processed across 16 institutions, some of which had not taken part in the first project. In all, 178 collections totaling 5865 linear feet were processed, across 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 two-part 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 look at 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

In many respects, the finding aids database has been an impressive success. There are currently about 2800 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 this spring, there are more than 900 finding aids from about 150 repositories represented in the database, and nearly all of these are collections that were invisible to the larger research community until the project and the database provided a way to catalogue 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 just a few months ago, and then the rest of the PACSCL institutions, with about 500. It is this last group that I want to look at more closely, since PACSCL is the group that set the project in motion, and its members are the bigger institutions in the region, with professional staff, extensive historical collections, and missions to serve the research needs of scholars, students and the public. Several PACSCL members have contributed

significant numbers of findings aids, particularly Drexel, Haverford, Bryn Mawr, the Free Library, the College of Physicians, and the Hagley Museum and Library, which is a special case and will be discussed separately. But most others have not. We had expected different results. 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 to it. 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, the ones that have been reporting), and 8 said that they were uncertain.

The reasons for the uncertainty and lack of continued submissions vary with the 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 in to 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 have had to deal with institutional IT policies that have blocked access to the harvester. There are also problems with the database itself that have 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 the institutions' websites moved and the harvester could no longer locate their finding aids. None of these problems is 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, and 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 were thinking 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 had been able to process 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. Other institutions had important unprocessed collections but they fell outside of the Philadelphia subject focus of the grant projects. Because these institutions had not been involved in the Hidden Collections grants, they had not given much consideration to the finding aids database, and in fact, several had not realized that contributing their finding 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 a few months ago, and its finding aids now constitute nearly a quarter of the finding aids there. 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, a site that is now up. Once the finding aids were in their own database, it was a relatively straightforward matter to make them available to the PACSCL site as well. At this point the benefits of having their finding aids in the PACSCL site isn't clear since 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 seeing 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 is not seen to be a good use of limited staff and resources. It is hard to argue with this point of view since 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 internally 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 nearly 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, but nonetheless, Penn is paying all of her salary, not PACSCL. 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 who Holly has been able to call on for assistance.

PACSCL has been extraordinarily successful over the last thirty 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 catalogue and process collections, something we have gotten very good at doing 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. Right 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 increasingly complex, and yet also 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. And 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 only be visible if they can partner with bigger players. Even the wealthier institutions are not always able to draw on IT support in the way that they would like. 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 the kind of 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 Penn, 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 hand-shake 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 treasures 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 people doing research. The Philadelphia region is full of situations where family papers are split among several institutions, or records of organizations that dealt with similar social, cultural and political issues are housed in different repositories, or documentation on

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 ambitious 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 Hidden Collections 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

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 this short 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.

Before doing so, I would like to preface my remarks by citing another example: The Digital Scriptorium. Unlike PACSCL, the Digital Scriptorium is a national consortium of academic libraries who have chosen to contribute selected digital images from each institution's collection of Medieval Manuscripts. There are currently thirty-six members of the consortium. What differentiates the Digital Scriptorium from PASCL is the fact that it does have an

institutional host: The University of California Berkeley Library. Without going into great detail, the Digital Scriptorium (DS) 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 Digital Scriptorium. 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 care and feeding 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 the 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 Berkeley. Thus the project which was launched at Berkeley with NEH funding was now returning to Berkeley as the technology host without funding and the curatorial center of the project remains at Columbia. Like other long term Digital Humanities, the DS is faced with long term technology maintenance issues: uploading new data, web development, database issues and data storage, among others. As technology host, Berkeley like other institutions, has off-loaded the cost of maintaining the DS site by charging the DS membership with a twenty-five thousand dollar annual fee. Based on a set of standard metrics, each member of the DS pays annual fee towards the upkeep of the DS database.

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

gifts or apply for grants. Thus the DS is in a state of limbo. Affiliated with two institutions and yet apart. PACSCL is in a similar situation; more, importantly the consortium, does not have 501 (c) 3 status and is not grounded within a specific institution. Given this situation, it is necessary for a PACSCL member to volunteer to serve as home institution for the grant. 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. UPenn Libraries was a founding member of the Philadelphia Special Collections Consortium. UPenn is, if I may say, modestly, among the largest 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 to see if UPenn might serve as institutional host both for the grant itself; and 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 and, in turn, the 2012 CLIR award. What does institutional support actually mean?

First it meant that Penn Libraries would serve as the 501 (c) 3 institution which 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

It should also be noted that the Project Managers for both grants were not only hired as University of Pennsylvania employees, (as were the graduate student processors), but they were also expected to report regularly to the PACSCL board regularly. In terms of project oversight

in the case of both CLIR grants there were two Principal Investigators, one representing PACSCL and the University of Pennsylvania Libraries.

It should be noted that PACSCL was very fortunate in hiring project two managers who were exemplary. Reminding them that PACSCL and Penn's reputation was dependent upon the successful completion of their respective projects, of course was never necessary.

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 Resource staff who assisted with the posting of jobs; dealing with payroll issues. This holds true for the Library's business office which actually managed the monetary grant and generated the financial reports for inclusion in our interim and final reports.

While finding offices in a building with few free office spaces for staff proved challenging but not insurmountable; the one issue for which Penn, (and I will confess that we did not participate in the grant project for processing our hidden collections), was technology. Apart from buying project laptops, setting up phones and computers or moving phones and computers (during CLIR I the project staff were relocated to another location in library due to construction here 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 twenty 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 were expected to manage site technology problems: setting up local instances of Archivist's Tool Kit, creating web folders, staff training and troubleshooting. It was part of their job description.

Now that we have completed the two CLIR PACSL hidden collections processing grants the lingering issue of Penn continuing to serve as host to the EAD PACSCL finding aids database is a question that has not been resolved. The data is secure. The finding aids are searchable and accessible. But like the Digital Scriptorium example I cited earlier, what is the cost of long term maintenance, not 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.