INTERNATIONAL DIGITAL LIBRARY RESEARCH & DEVELOPMENT MEETING

STANFORD UNIVERSITY, NOV. 29-DEC. 1, 2009

Summary report prepared by
Mimi Calter, Stanford University
Elliott Shore, Bryn Mawr College
Christa Williford, Council on Library and Information Resources

Introduction
In November 2009 Stanford University and the Council on Library and Information Resources (CLIR) co-hosted an international gathering to discuss opportunities for collaboration in digital library research and development. The sixty-one participants, who represented thirteen countries and twenty-four institutions and included both senior administrators and senior technologists, came away energized and engaged to an extent that many found unexpected. *"I experienced...surprise at the articulation of our shared issues,"* said one. Others commented, "*I leave with a lot more work to do!" "I am overwhelmed -- thrilled with the outcomes", "Interesting and exciting -- This dream might come true."* These comments are remarkable in light of the difficult decisions these leaders have been forced to make during the global financial crisis; in an era of retrenchment, they are seeking to extend the reach of their institutions through unprecedented integration of their digital collections with others' in the international library community. As sponsors, Stanford and CLIR are gratified by this level of enthusiasm, and look forward to continuing such discussions and facilitating collaborative projects at the global level. This report strives to encapsulate the process that inspired such hope for the future, and begins to look at the next steps required to advance on the journey toward greater cooperation.

Several factors contributed to the group's sense of engagement and *bonhomie.* Many participants were previously acquainted with one another and so were already familiar with the common ground that they shared. The European national library directors who attended have been working closely together on digital library projects for many years, while the Japanese researchers and developers present at the meeting have together cultivated a deep and nuanced understanding of technical issues affecting international cyberinfrastructure. Recent activities led by CLIR and the Digital Library Federation (DLF), now one unified organization, have focused on whetting the participating Americans' appetites for collaborative digital library initiatives within the context of CLIR's larger mission of helping library leaders create and maintain stronger links between academic libraries and the scholarly disciplines. In addition, the Stanford University setting was spectacular: sunny weather, beautifully prepared and interesting meals, and a tight schedule of varied

activities designed by the meeting's organizers kept participants alert and focused on the work at hand.

But the common history among sub-groups of participants and the positive environment in which they met were only a part of the story. Larger, more significant factors were a recognition of shared concerns, vision and purpose, and a growing sense of frustration at how long it has taken for work on a truly integrated international library cyberinfrastructure to begin: *"We actually need one another--it is urgent,"* said one; "*It is a most exciting and most terrifying time--we have such a big responsibility--we need to be positively impatient,"* stated another. The group came away with an understanding that consisted of two linked realizations:

1. The piece-by-piece maturation of the digital library, growing as it has from separately funded initiatives that are often redundant, is neither efficient nor sustainable. Coordination and collaboration are no longer merely good ideas, they are essential for the survival of our cultural heritage. Furthermore, the distinctions between public, academic, and national libraries and archives are blurring, and their user communities are increasingly diverse and overlap one another.
2. The idea of a user-centered coherent digital library seems possible in a way that it never has before, but the challenges of collaboration on a global scale make it difficult to imagine how such an environment might be built or managed. Striving to achieve widespread agreement on priorities for digital library development would necessarily be a lengthy process that could slow down valuable work, or even ultimately be counter-productive. Throughout the meeting, there remained an apparent opposition between "big picture" visionaries seeking consensus on long-term goals and those wanting a more practical, step-by-step, approach to tackling specific, immediate needs. The tension between these perspectives was never resolved. For this reason, rather than trying to reach a unified agenda, the group sought to identify opportunities for subsets of institutions to collaborate in both conceptual and practical ways.

In order to represent the richness of the meeting's discussions, the report consists of two major sections: first, a summary of the themes that emerged as the group sought to conceptualize their shared problems, and second, a list of collaborative projects that participants noted as promising steps forward for international digital libraries. The group focused on identifying the characteristics of (1) sustainable and extensible partnerships, consortia, and subject and regional domains of excellence, (2) a genuinely robust digital environment for teaching and research, and (3) discrete initiatives that might contribute to the achievement of both. The collaborative projects identified by the group--some new, some ongoing, some aspirational--indicate the level of enthusiasm for the job ahead. The report concludes with some remarks about next steps to be taken to further the goals expressed at the meeting.

Themes

The "big picture" and practical questions noted above seem to reside in a wonderful and fertile tension to one another, a tension arising from a number of sources: desiring to do things well for one's institution while doing things well for the larger public good, needing to work closely with others while retaining a distinct institutional identity, and determining what is the minimum level of consensus necessary for fruitful collaboration. Should one focus on large goals or on the small, incremental steps required to reach them? Must all collaborating partners do their work the same way, or might they work together without a fixed, agreed-upon set of policies and procedures?

Such questions complicated even the earliest discussions at the meeting, at which organizers presented a detailed taxonomy of digital library functions and services drawn from hundreds of pages of background reading about initiatives and programs contributed by participating institutions. Looking at the complexity and interrelatedness of the concepts on this "map" of the digital library "landscape," the group could see that the boundaries between the terrains of search and discovery, storage and preservation, and collection development and maintenance are not often clearly drawn. It soon became clear that a "divide and conquer" plan of attack on such a landscape, enlisting separate institutions with the responsibility of covering separate "flanks," would not be workable. The taxonomy in some ways replicated the ways that libraries have traditionally conceptualized their work, as autonomous performers of a more or less generally agreed upon set of tasks, collecting and organizing materials, preserving and presenting them to their publics. The digital era seems to require a radical rethinking of this conceptual vision, one that needs to recognize the need for interconnectedness among libraries and those processes that seemed once to be discrete from one another.

John Wilkin of the University of Michigan offered a way to reconceptualize the connectedness required for successful institutional collaborations. When partners view the issues they face as "*common problems*," in other words, typical or commonly seen challenges that each partner must address, the tendency is to define those challenges within each institutional context and to begin to work on solutions separately before engaging with others. While the "*common problem*" approach may be mutually beneficial, it reinforces institutional boundaries and the tendency to work in isolation. Wilkin encouraged the group to instead see their issues as *"shared problems*," requiring input from all partners for resolution. *"Shared problems"* require participants to pool resources and work in a collective space. Chuck Henry of CLIR, in reviewing Wilkin's remarks, concluded that how a group frames a problem philosophically determines its members' ability to work together to resolve it. Wilkin cited the development of institutional repositories as a classic case of trying to address a commonplace problem: institutions share code and marketing tips, but work separately to implement their own software instances and storage strategies. He argued that there are remarkably few examples of efforts directed at "*shared"* problems, which would require moving beyond institutional and national

boundaries, making long-term commitments to collective solutions that may not perfectly match any individual partner's priorities.

Explained in this fashion, the paucity of initiatives addressing "*shared problems*" is hardly surprising, since such efforts bring with them a high degree of risk. But the question remains: in today's technological and economic climate, is risk aversion an acceptable deterrent to deep collaboration? Participants in the meeting seemed to think not: "*The challenges are international and the solutions need to be as well,*" said one; "*We are progressing along a spectrum of more sharedness, more shared approaches,*" observed another; a third stated, "*Cooperation, coordination **are** our shared values.*" Still others remarked that, whether or not intended to be so, *"digital libraries **are** global today*," and the wise choice is to acknowledge fully the contributions of our international colleagues and build upon rather than replicate, them. After all, for the user of a digital library or archive, institutional and national boundaries are much less significant than they are for the librarian or archivist. This is not to say that organizational and national cultures are not important; indeed, they are critical factors in determining access to resources, working conditions, professionalization, or funding. The point is that because the community of users of digital libraries is international, it no longer makes sense for library and archival professionals to preserve cultural memory without reference to the complex and rapidly changing international context in which the artifacts of that memory will live on.

Wilkin cited several recent movements toward a shared digital library cyberinfrastructure that showed the promise of becoming global, namely, the Blacklight discovery interface, the Duraspace repository technologies, and the Hydra Project, which marries Blacklight discovery with the Fedora repository. While praiseworthy in that they enhance and may ultimately bring together the best quality and most powerful functions of the online search experience, he noted that each of these tools may still be implemented separately, institution by institution. By contrast, multidisciplinary, multi-institutional projects like Ethicshare may cross institutional and national boundaries, but in order to flourish, they need to continue to grow. Wilkin challenged the group to more ambitious forms of collaboration. As discussion continued, many participants asserted that forging connections outside the domains of cultural heritage and education will also be important; in other words, it behooves libraries not to look with anxiety at commercial information providers and tools as rivals, but rather to assume fully the responsibility to preserve and disseminate the cultural heritage of humankind while taking advantage of any added value businesses such as Google, for example, can provide.

The format of the meeting emphasized small group interactions, including structured visualization exercises as well as both formal and informal conversations. Participants stressed the importance of the variety of these face-to-face interactions to the experience; one observed, "*sometimes you have to cross the ocean to meet people who are right next door.*" The length and types of interactions had been designed by Stanford staff in cooperation with professional meeting

facilitators Andrea Saveri and Eris Weaver, but the questions explored through these activities were generated entirely by participants. One of their major concerns was the seemingly diminishing role that libraries play in research activities; participants phrased the question as follows: "*From the user's perspective, how might libraries fit into the 'digital value chain'?*" For example, attendees felt that scholarly users of today's libraries view themselves as contributors to disciplinary discourses rather than as generators of a single institutional legacy. For this reason, libraries need to foreground their own contributions to disciplinary development beyond the confines of the institutions at which they are based, and to make the significance of these contributions clear in an global context. Rather than striving to meet all types of information needs represented in their local user communities, they can instead strive for excellence in their areas of strength while at the same time building trusted relationships with other libraries to serve the other needs of their constituencies. Competition among libraries for collections and among commercial entities who sell information products and tools will remain, but librarians must become conscious of the ways such competition does not serve the interests of users; they must think strategically and creatively about new models of ownership when these can improve access to resources and the quality of services. An important step toward international collaboration could take the shape of an organized effort to argue the case for viewing and supporting the global digital library as a public good. Chuck Henry of CLIR has agreed to lead such an effort by publishing a series of essays on this topic.

A second, related concern was summarized in the question, "*How might we encourage sharing and distributing risks and responsibilities?*" Here the group saw the necessity of attacking the question in both visionary and practical ways. Participants agreed that there are many library functions that might be successfully distributed among institutions--such as collection development, off-site storage, digitization, or specialized conservation and preservation services; however, building the inter-institutional trust necessary to de-duplicate such functions can be very costly, both financially and politically. Building loose partnerships with like institutions that strengthen gradually has been one time-tested approach. Other participants stressed the need for an honest initial assessment of each partner's strengths and weaknesses, followed by the establishment of clear objectives and realistic timelines that set out potential risks, rewards, and measures of success. The group noted that outsourcing tasks to goups of institutions that no individual partner can handle comfortably may become necessary, and they stressed the importance of encouraging partners to improve in their areas of specialization while gradually increasing reliance on their expertise. Another observation was that long-term success might require accepting that partner contributions may sometimes be unequal. Finally, participants acknowledged the need to explore openly past failed collaborations between cultural heritage institutions. Issues affecting both successful and unsuccessful collaborations have been highlighted in recent reports such as "[Beyond the Silos of the LAMS: Collaboration Among Libraries, Archives, and Museums](#)" (Zorich, Waibel, and Erway, OCLC Research, 2008); further work specific to international partnerships and consortia would also be welcome.

A third concern of participants was the balance between specialization and generalization, underlying the more specific questions: "*How might libraries build and operate a discovery environment across silos?*"; "*How might libraries promote collective curation?*"; and "*How might libraries collaboratively approach digital preservation?*" In a discussion of the first of these questions, rather than lamenting the "silo-ification" of digital library resources as an unhealthy and wasteful trend, participants cautioned that the fear of closed information systems might be overblown. Instead, in certain contexts silos can be productive for the advancement of knowledge in a multilingual, multicultural world. Rather than focusing on changing legacy systems, the group advocated that international libraries support diversity among systems and strive to expand the number of ways to search across them. In addition, they suggested that digital library managers encourage a moderate amount of redundancy in content and tolerance for idiosyncratic approaches to description. Similar ideas were expressed in the discussion of collective curation, where participants envisioned library partners staking out discrete yet overlapping curatorial domains in conversation with partner institutions. The goal of a curatorial partnership would be to build disparate collections of digital objects while sharing them across loosely federated institutions. The balance of centralization with decentralization in collaborative efforts also figured in the related discussion of shared digital preservation initiatives. Here the consensus seemed to be that strong agreement around shared responsibilities and priorities was necessary while maintaining a diversity of separate, local storage environments.

As the threads of the discussion looped back to the abstract concepts around which notions of the future of the digital library are based, participants challenged themselves to express their visions of that future in a number of different ways. In particular, a group focusing on the question, "*What will the business of libraries be in 5 to 10 years?*," struggled with whether such a vision might be shared, and, if so, whether it should be described as reflecting revolutionary, or evolutionary change. While the question of evolution versus revolution remained unresolved, participants seemed to agree on this point: the basic mission of libraries will remain intact, but the roles libraries are playing as preservers and cultivators of knowledge are becoming more challenging to define. More than one participant observed that library services are expanding in scope from one discrete area of the "information lifecycle"--i.e. the collection of published materials--to encompassing the whole environment in which the consumption and production of knowledge takes place, including exchanges between teachers and learners, information objects and users, and scholars with their peers. Participants seemed to feel that this more diffuse environment is as exciting as it is challenging, not least because few parts of this environment now fall into the exclusive purview of institutions that call themselves libraries. Still others tried to express the changing nature of libraries from the standpoint of their collections: as the proportion of analog to digital holdings shifts in favor of the digital, the potential to serve greater numbers of people who do not visit a particular library building increases along with the difficulty of predicting and

understanding their needs. Greater selectivity and a greater investment in preservation will be required. Because it is time-consuming and expensive to do either collection development or preservation well, sharing these responsibilities across institutions and making a conscious effort to promote the contributions of cultural heritage organizations to global civilization will be requisite of all future library leaders.

Collaborations
A key piece of the closing day, and a key goal of the meeting, was to raise awareness of existing collaborations between participants, and issue invitations to new potential partners. A summary of the initiatives discussed at the conference is below. They vary greatly. Some are ongoing collaborations, with the potential to integrate new organizations into a group; others are newly formed collaborations, based on shared problems identified at the conference; and a few are aspirational-- projects or initiatives envisioned or undertaken by a single institution, but with opportunity for further interactivity and partnership. Some projects are focused on a specific tool or technology while others look more holistically at a problem set. In many cases, the number of institutions rather than the institutions themselves are named, in order that directors can review their commitments to these projects with their staff before publicizing their participation.

The meeting was designed to nurture relationships, and to allow partnerships to form organically. The expectation is that each collaborative group will take its work in the direction most appropriate for its members. At the same time, participants stressed the need to continue open discussions of collaborative efforts; at a very basic level, promoting general awareness of worldwide research and development activities is the first step toward a globally integrated system. Furthermore, digital library developers need to be aware not just of the successes but also of the failures that colleagues have experienced, in order to avoid making the same mistakes. In future exchanges, the group will revise and expand this list of collaborations, critically assess their progress, and identify future steps that might forge further connections between initiatives.

<u>Newly formed collaborations</u>
1) Repository Architectural Review
There were a variety of point-to-point conversations and one small group discussion about revising local repository design, and the need to establish a best practice or reference architecture for repositories. The small collaborations that have been identified include:
- A consultative review of Stanford's digital library architecture by another participant
- Two investigations of existing digital library implementations between partner participants

- The creation of a small group to identify a reference architecture for digital libraries & repositories. It is anticipated that this team will coordinate with Sun's PASIG Repository Working Group, and a presentation at the next PASIG meeting is planned.

2) Data Curation and Preservation
Three participating institutions, including Stanford, have agreed to coordinate efforts on data curation and preservation, and will seek out a limited number of partners for an initial collaboration. Participants from Stanford have submitted an abstract on the topic for the IATUL (http://www.iatul.org/) eScience meeting this summer at Purdue, and believe that Purdue, though not a conference participant, has a strong interest in this area. Chuck Henry from CLIR offered funding to help organize a data curation effort across a group of US libraries.

3) Digital Manuscripts
Several participants are active in the area of digital manuscripts, and have discussed coordinating efforts. The British Library is interested in reusable software components for manuscript digitization & delivery. The Danish Royal Library is committed to contributing 100 medieval manuscripts to the digital manuscript commons (https://wiki.library.jhu.edu/display/MSSCOMMONS/Home), with a shared infrastructure, preservation and sustainability plan. CLIR also expressed a strong interest in digital manuscript commons, particularly with regard to the sustainability of the effort.

4) Digital Forensics
Three participating institutions, including Stanford, have begun talks on collaboratively developing a digital forensics program and toolkit.

5) Semantic Web
There is general interest in this emerging technology. The British Library has agreed to host a conference on the topic in 2010.

Expansion of ongoing collaborations & projects

6) Hydra / Blacklight
Hydra (http://www.fedora-commons.org/confluence/display/hydra/The+Hydra+Project) and Blacklight (http://projectblacklight.org/) are being developed as tools for building an environment for hosting and managing digital assets. Blacklight is the discovery mechanism. Stanford, the University of Virginia, the University of Hull and Fedora Commons have been working in partnership on this effort, have had some initial success, and are looking to expand. Eleven potential partners were identified at the conference.

7) PLANETS

Lynne Brindley and Sean Martin from the British Library encourage participants to adopt and sustain PLANETS (http://www.planets-project.eu/) once it moves to an open source software foundation. This five-year EU project has been successful and productive in building tools, codifying best practices, and fostering a community around digital preservation in Europe.

8) GETA

Participants from the Nippon Institute for Informatics took time after the conference to review Stanford's plans to integrate GETA (http://geta.ex.nii.ac.jp/e/index.html) into the Google Books workflow and SearchWorks, and helped install the latest version of GETA locally.

9) Hathi Trust

The trend toward the consolidation of libraries in the next five to ten is probably inevitable. Hathi Trust (http://www.hathitrust.org/) would like to see more institutions join and expand the organization's shared repository, taking advantage of the opportunity to deal with print management in a collaborative way. Participants from Hathi Trust encouraged those with digital content assets to contact them.

10) Scholar's Workbench

Duraspace's Scholar's Workbench community (https://fedora-commons.org/confluence/display/FCCWG/Scholars+Workbench) is already international, and is particularly active in the United Kingdom. The program has room for additional members from around the world, and welcomes further participation.

11) DRIVER/DNET

DRIVER (http://www.driver-repository.eu/), Digital Repository Infrastructure Vision for European Research, is an open-source network of institutional repositories in Europe, Australia, and Japan. DRIVER is a multi-phase effort whose vision and primary objective is to create a cohesive, robust and flexible, pan-European infrastructure for digital repositories, offering sophisticated services and functionalities for researchers, administrators and the general public. All participants are invited to become involved with the network. The D-Net software package supports the DRIVER network.

12) Europeana

Europeana.eu is a collaboration between universities, research institutes and content providers to test search tools and provide access to digital resources. The website is now in its beta testing form, and conference participants are strongly encouraged to test, use, and offer opinions on the site and on Europeana services.

13) National Digital Library of China

The [National Library of China/National Digital Library of China](http://www.nlc.gov.cn/en/indexen.htm) is seeking collaborative partners as it seeks to build the largest digital library in the world. Together with the Library of Congress, the National Digital Library contributes to the [World Digital Library](). The National Digital Library would especially like to share catalog records and metadata for digital collections.

14) BIBAPP
The University of Illinois Urbana-Champaign and the University of Wisconsin have developed [BibApp](http://bibapp.org/) (http://bibapp.org/), a research-tracking tool that mines the web for current publication data for communities of researchers. A free pilot version of the tool will be offered in February 2010, and the developers are seeking partners to test and adapt the software at other institutions.

15) STITCH@CATCH
Part of the Dutch scientific research program [CATCH](http://www.nwo.nl/nwohome.nsf/pages/NWOP_5XSKYG_Eng) (Continuous Access To Cultural Heritage, http://www.nwo.nl/nwohome.nsf/pages/NWOP_5XSKYG_Eng), STITCH (SemanTic Interoperability To access Cultural Heritage, http://www.cs.vu.nl/STITCH/) is an initiative focused on building cross-search capabilities for digital library collections. The initiative seeks to expand its range of use cases and testing partners.

16) SHAMAN
SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg, http://shaman-ip.eu/shaman/) is a primarily EU-based project to create an environment for managing the storage, access, presentation, and manipulation of digital objects over time. Participants report that SHAMAN needs worldwide partners to test its prototype applications for archiving scientific research papers and data, industrial design and engineering papers and data, and parliamentary documents.

Aspirational projects

17) Text Mining
The University of Virginia is seeking partners to develop text-mining services to be applied to the Google corpus.

18) Organizational and Development Resources
Emory and Stanford universities would like to work together on improving and promulgating library management and professional development resources, including organizational staffing models, processes for the evolution of these models over time, and workflows for digital library development projects.

19) API Development

The Bibliothèque nationale de France would like to collaborate on the development of application programming interface standards for digital library systems.

Next Steps

We have a strong sense of optimism," observed one participant at the last session at Stanford, "*[We have been] reaffirming a shared vision.*" Although largely satisfied with the three-day conference, many left concerned about how they might best translate the sense of urgency they felt to the staff of their home institutions. Communicating the significance of global cooperation to the future of cultural memory organizations is now the responsibility of all who attended, one of whom noted that, if successful in their efforts, the day-to-day contributions of their staff will begin to take on "*that much more importance.*"

While persuading others to join their efforts is a key first step, the true measure of the success of events described in this report will be in the progress of the collaborative initiatives identified above, and in the other efforts engendered through succeeding conversations between participants and their other international colleagues. At the close of the three-day event, there was a widespread acknowledgment that this would be a large task: *"I have a renewed sense of privilege and responsibility as a leader,*" observed one participant. "*It is a most exciting and most terrifying time and such a big responsibility that we face*," said another. Enthusiasm for scheduling additional conferences in the near future was high, especially if such events facilitated very open discussion of the urgent problems facing libraries and the challenges of building successful collaborations, in the words of one participant, "*not just sharing success stories.*" A date for such a meeting has not yet been set, but initial talks have focused on a possible meeting in Europe in 2010.

Reflecting on their original intentions for organizing the initial gathering at Stanford, Mike Keller of Stanford and Chuck Henry of CLIR set out three key goals:

1. That the group should develop a shared understanding and taxonomy of key areas of concern for digital library research and development, and a commitment to sharing both the positive and the less successful results of initiatives linked to these areas with the larger community of institutional adopters of digital technologies;
2. That they should use this shared understanding and taxonomy to collectively identify gaps of functionality, systems, and the operational environments of digital libraries;
3. That they should establish small working partnerships or collaborations among their institutions committed to developing applications that would fill such gaps.

Keller and Henry believe some progress was made toward each of these overlapping goals, but that work remains in identifying gaps of functionality, in systems, and for operations. In order to continue to nurture and celebrate activities arising from the working collaborations identified above, staff at CLIR-DLF will allocate a portion the [organization's website](#) for collecting and providing further information about these projects, as well as expanding this list to include other opportunities for libraries seeking collaborative partners. Leaders of each project will be asked to expand upon and clarify the goals of each initiative, and to update CLIR as projects progress. Rather than building this web resource as a simple inventory of digital projects, CLIR staff hope that the community of leaders engaged in these projects will be able to use the resource to continue the development and expansion of the shared understanding and nascent taxonomy begun at Stanford. Finally, beginning with a series of essays on the subject of promoting the global digital library as a public good, CLIR and the Digital Library Federation (DLF) will take responsibility for commissioning further research to promote global awareness of issues affecting international collaboration for libraries and other cultural heritage organizations. CLIR, DLF, and Stanford will work with project and library leaders to continue to build widespread support for the cause of global cooperation for cyberinfrastructure and to publicize opportunities for those who wish to participate in its advancement.