DLF QUARTERLY UPDATE:

JULY 1 –SEPTEMBER 30, 2004

CONTENTS

SUMMARY

The first quarter of the year, with its dog days of summer, is typically a quieter period for travel, presentations, and meetings – at least until the semester kicks off and September rolls around. And so it was for this quarter, although the June **Scholars Panel** meeting and its aftermath, the planning for the early August DODL meeting – during which it transformed into **Aquifer** – and the program setup and fellowship competition for the Fall Forum (Baltimore, October 2004) more than filled the time. The Aquifer Group has set itself a series of commitments and goals with a firm timeline, and I worked steadily on the communications opportunities for us (including a foray into **wikis**) and the design of a **pair of registries** for existing digital library tools and assessment evaluations, both areas where the group felt coordination and potential joint development could occur.

The quarter saw several additional highlights – the arrival of **Nancy Davenport** as the permanent CLIR president took us out of a year-long interim, and we are a much cheerier office now; we won our first (I think) competitive federal grant as an organization, with the happy news at the end of the period that we had been awarded a $292,000 **IMLS National Leadership Grant** for OAI Best Practices work**;** trips to Brighton, England and to the New Zealand Library Association conference allowed me to cement our relationship with **JISC**, now our newest ally, and to begin a series of dialogues with various New Zealand institutions – a country rich in digital library achievements and national ambitions; we produced two new online publications: *Electronic Resource Management,* the culmination of two years of work by a DLF team led by Tim Jewell (Washington) to articulate a common XML-based record to contain license terms, which is already seeing real interest from publishers, content aggregators, and from library system vendors, all of whom see the potential value and cost savings in a common way to articulate, exchange, and manage licenses to subscription databases and journals; and *Digital Library Content and Course Management Systems: Issues of Interoperation* – a DLF study team funded by the Andrew Mellon Foundation and led by Dale Flecker (Harvard) and Neil McLean (IMS Australia) that maps out the next steps for us in the critically important area of library content and courseware integration. I'm sure this study will be the bedrock for much future work by us and others – the area is one that is of importance to us all as we struggle to get maximum teaching value out of the digital library materials we build and buy.

The following report is in three parts, a pattern I will maintain in future Quarterly Reports unless feedback dictates otherwise: a calendar of events, to give a sense of the range of my activities; a feature on a particular initiative – this time the findings from our recent Scholars Panel; and an Executive Summary on a particular technology or standard (this time, Wikis).

Looking forward to the busy months to come.

David

DLF CALENDAR OF EVENTS AND ACTIVITIES: JULY-SEPTEMBER 2004

7/6 Nancy Davenport starts at CLIR

7/6 Presentation: "Institutional Repositories: An Overview." JISC Program Meeting. Brighton, England.

7/8-7/10 Attended JISC/CNI Joint conference. Brighton, England

7/12 Yale site visit and update: meetings with Alice Prochaska, Meg Bellinger, Jeff Barnett, Dan Chudnov, Fred Martz, Audrey Novak, Sandra Peterson, Ann Green, and Julie Linden. Discussed implementing Endeavor's Voyager integrated library management system; DODL; American Studies Digital Imaging Project; Virtual Reference Pilot Project; Shoah Foundation video delivery.

7/13: The Economic Growth Center Digital Library External Review Committee. Sterling Memorial Library, Yale. [Judy Russell, U.S. Government Printing Office; Mary Vardigan, ICPSR; Steve Puglia, National Archives and Records Administration; Mark Maynard, Roper Center; David Seaman, Digital Library Federation; Dan Hazen, Harvard University].

7/14 Meet in Washington DC with Sarah Porter, incoming Head of Development, JISC

7/21 Meet in Washington DC with Liz Bischoff (OCLC)

7/23 Meet in Charlottesville, VA, with the scholars from the Virtual Jamestown board, to follow up on Scholars Panel discussions

7/29 OAI Best Practices Meeting, CDL, Oakland, CA [Amy Harbur attended in my stead]

DLF publishes *Digital Library Content and Course Management Systems: Issues of Interoperation* (July 2004) http://www.diglib.org/pubs/cmsdl0407/

DLF publishes *Electronic Resource Management* (August 2004) http://www.diglib.org/pubs/dlfermi0408/

8/5 DODL meeting, Stanford University, Stanford, CA – now Aquifer

8/16 CLIR annual report -- Digital Libraries section: http://www.clir.org/pubs/annual/annrpt2003/03annrep.pdf

Work with Eric Celeste (Minnesota) on Aquifer webpage, wiki, and communications plan. See www.diglib.org/aquifer/ Work with Michael Keller and Dan Greenstein on the job description and vetting of candidates for the Aquifer Director's position.

Planning for DLF Fall Forum – program committee; Forum Fellows selection.

8/20 Planning for CLIR/DLF "Managing Digital Assets" meeting in Charleston, SC, Feburary 2005.

8/30 Wrote and distributed to Aquifer group *A Quicki Guide to Wiki* as part of the Aquifer communications investigation. [See *Executive Summary* below]

9/1 Welcomed David Ferriero (NYPL) to DLF – set up visit to brief him on DLF in person in October.

9/7 Keynote presentation: Lianza – New Zealand Library Association, Auckland: "Mass, Malleability, and the Collaboration Imperative: Trends for the Digital Library." [To be given again at the Australian National University, as the 2004 James Bennett lecture in early November. Previous lecturers include our own Michael Keller and Lynne Brindley.]

9/8 Led the New Zealand IT Special Interest Group Workshop on discussion of repositories and harvestable metadata

9/8 Attended *Matapihi* launch – NZ art collection online, Auckland Art Gallery. Part of a grander *NZ Online* national ambition, including recent NZ$24 million government contract to NZ National Library for a national repository.

9/9 Meeting with University of Auckland librarian Janet Copsey and Brian Flaherty, Digital Services Manager: discussed Auckland's plans for an institutional repository

9/10 Meeting with New Zealand National Librarian Penny Carnaby; presentation to New Zealand National Library digital team.

9/14 presentation in Washington DC to visiting librarians from Serbia and Montenegro

9/14 Meeting with Sam Black of Squire, Sanders & Dempsey LLP to discuss DLF legal independence issues

9/22 DLF update for CLIR staff (lunchtime session, provided periodically, to keep CLIR staff informed about DLF initiatives)

9/23 DLF wins IMLS National Leadership Grant for OAI Best Practices work.

FEATURE ARTICLE: DLF Scholars' Panel

Introduction

College and university librarians have a long tradition of listening to their users, and of adjusting our services and collections according to the articulated needs of our faculty and students. This is nowhere more important that in our emerging digital library endeavors, where much is still unfamiliar to patrons and where new products, aggregations, tools, and services come (and often go) with confounding frequency. Across the many initiatives, benchmarks, and standards undertaken by the Digital Library Federation (DLF) is an overarching desire to build library services and online holdings that result in richer scholarship and more effective pedagogy.

To help inform this end, and to test our own assumptions about developing needs in digital scholarship, the DLF convened in the summer of 2004 in Washington, DC, a group of humanities and social science practitioners, all of whom are actively building digital archives, online editions, and electronic scholarship to further their academic and teaching interests, and who are working with their library colleagues and digital collections in innovative ways. Over two days of lively and free-flowing discussion in June, the scholars provided feedback on how libraries could partner with them to serve their particular digital scholarship needs [a bibliography, list of participants, and list of topics can be found online at http://www.diglib.org/use/scholars0406/ or http://purl.oclc.org/dlf/scholars0406/]. Follow-up discussions with several participants have further fleshed out the themes and observations detailed below:

Barriers to Digital Scholarship

We turned our attention early to the hurdles that face this first wave of scholars undertaking serious digital scholarship, in order to understand which of these barriers can be overcome by emerging digital library research. There was speedy and widespread agreement that an overarching problem was the lack of persistent identifiers – permanent and trusted internet addresses – for online objects. How can you invest in rich, hyperlinked scholarly writing or scholar-driven archives if the material not under your immediate control keeps moving from web address to web address, or disappearing altogether (an irritation commonly known as "link rot")? It is a waste of time to have to monitor and fix broken links, and a disincentive to undertaking further work. This is a problem they look to libraries and publishers to solve, and to solve quickly ("aren't you guys supposed to be good at this sort of standardization" one participant said). As a positive example of persistent identification in the scholarly journals industry we looked at Crossref and the Crossref/Google article search service, which have grown up around the Digital Object Identifier (DOI) persistent ID that is commonly used in the STM scholarly journal arena.

The other main thread of this conversation about hindrances to digital scholarship – which came up in discussions of institutional repositories too – was the failure of departmental promotion and rewards structures to recognize and accommodate the shift

from a print-based to a digital world of scholarly publishing and communications. It is no accident that most active humanists and social scientists working with digital media are post-tenure, one participant observed, and I suspect that even then they are not all immune from the career-depressing effects of being seen to be "too digital" or "only digital."

Need for Tools

The group was clear that there is a severe need for tools customized for a range of scholarly inquiry needs:

- Gathering information from multiple sources, along with some information about it (personal libraries with metadata)
- Searching of images
- Visualization of patterns and trends and search results
- Annotation of text, image, and multimedia files
- Writing the new scholarship – authoring tools for the digital scholar

However, so unfamiliar is this area that we heard from several individuals that they had a hard time articulating precisely what they required from such tools, or what level of software creation skills or consultancy is available to them, and where. We are still in a stage where it is easier to react to an example of an existing tool than to dream them up *ex nihilo*, and with that in mind we discussed and demonstrated a variety of software packages that allowed scholars to gather, search, annotate, and re-package digital objects from library collections, including New Zealand's impressive *Greenstone* (referred to in this context as a personal library organizer), the suite of tools from UC Berkeley's *Scholar's Box* initiative, and Michigan State's *Matrix* annotation software that is aimed at various streaming media. Clearly a first-order need for this group was simply to know how to discover that these sorts of products exist (let alone the range of locally created but re-usable software custom-built for various initiatives), and what their characteristics are.

Services: Repositories and Harvestable Metadata

There has been a rapid growth in the ambitions of universities to build systems to safeguard and re-use the full range of scholarly and pedagogical output – the institutional repository movement. Opinions about this phenomena may well differ across disciplines; for this group there was a decidedly cool reception to the notion of turning over their scholarship, datasets, and archives to their institution for exploitation as institutional assets (the language of the institutional repository discussions may well be to its detriment – faculty do not necessarily take kindly to being cast as asset workers producing exploitable product for their institutions, even if only at the level of language).

While the ability to have a long-term safe-haven for their digital content found some real favor, especially as it was curated by the library, there was a range of concerns beyond this – questions of ownership, permissions, load (how much work is it to prepare a body

of material for a repository?) and again the observation that there was no link between the re-use of a scholarly asset and the current faculty rewards system.

Much more positive was the reaction to sharable and harvestable metadata – not a concept that was very clear to the group prior to the meeting. We used the Open Archives Initiative (OAI) as an example of simple metadata records for digital objects that are put on the web and harvested by software, in order to build services that include records from many sites all arranged in one service of portal. There was a good deal of interest in this mechanism both as a way to help make their own scholarship more visible, and as a way of gathering up references to related material to which they may want to refer.

Digital Library Collections

Given the active involvement these scholars have in building and contextualizing content – in engaging actively in the creation of digital archives that they then manipulate – and given the concern with link rot – it was no surprise to learn that a behavior they wanted from collections of digital objects was the ability to capture and re-use that material in their own local contexts. There was firm agreement that it is not always enough to link to a resource in someone else's system, even if the link is persistent. The need for a local copy may be aesthetic integration into an archive; offline use; incorporation into a desktop tool of some sort (data visualizer; annotation tool; courseware package; textual analysis software); data enrichment with terminology of the scholar's choosing; or even the simple need to search a body of material all at once – impossible when the books are in different systems with different search tools. Equally clear is how difficult it is to get permission from data holders to satisfy this common need, even when the material in question is freely available on the internet in archives and libraries. Typically the institutions who digitized and who host the material do not have policies in place, or rights expressions, to allow that content to have a secondary life in an online project at another institution. "Just link to it" is not the answer often for this group of scholars, but absent a mechanism to explicitly accommodate the desire to bring digital objects into a local scholar's archive, they are left with a frustrating and time-consuming series of conversations, favors, and personal pleas in order to engage deeply and actively with the material in digital library collections.

Closing

Work with this group has been lively and enlightening – for individual projects and in an ad hoc manner for the organization as a whole. Such scholarly users make for very effective reaction and review panels. After the event, several members articulated a need for help in acquiring either digital copies of items as yet undigitized or the permission to move digital items held elsewhere into their own archives and tools. The latter may well give us a clearer sense of how and when simple access is not enough, and close engagement with and enrichment of a file in another library's collection is what is needed to fulfill a scholarly or pedagogic need. In addition, one specific opportunity for partnership was put forward, by the Virtual Jamestown group; I enclose this as an Appendix in case it touches a nerve with any DLF library.

EXECUTIVE SUMMARY: Wikis

> *A wiki is a website that any user can edit – live on the webpage, in real time – as well as read. In its purest sense this means that any visitor to the site is also an author, with no technical controls on what they add or delete and no gate-keeping editorial process prior to posting; in more limited forms it means any user within a defined group who has access to the wiki – a project team working on collaborative documents, for example.*

A website is typically a read-only medium – you visit it, read it, print it, save it, but you cannot alter the content. In order to edit a webpage you need to be its creator, its owner. This is a normal, commonsensical arrangement – who would want anyone on the web coming to their site and making changes? Well, Wiki would.

> *Listen to this brief August 2003 National Public Radio feature for a clear explication of the shape, form, and surprising success of the wiki.*

Wikis are communal efforts, based on social regulation rather than technical safeguards – when vandals attack and make destructive changes, the much larger number of honest users fix the situation (changes are tracked and viewable publicly, so they can therefore be undone quickly). Vandal-users tire of doing bad because their damage is quickly undone by steward-users. Genuine mistakes by one writer can be fixed by another, to everyone's benefit; pieces of helpful knowledge can very easily be added by any visiting expert. Sounds unworkable, but there are some situations in which it is proving to be less fragile than it sounds.

The Wikipedia: this reference work **http://en.wikipedia.org**/ is the most famous and effective of the wikis you are likely to find on the open web. It is a 330,000 entry encyclopedia built by 100,000 contributors. Here's its definition of a wiki:

> *A wiki (pronounced "wicky" or "weeky") is a website ... that gives users the ability to add content, as on an Internet forum, but also allows that content to be edited by other users. The term can also refer to the collaborative software used to create such a website…. Wiki wiki comes from the Hawaiian term for 'quick' or 'super-fast'* (available from **http://en.wikipedia.org/wiki/Wiki**).

Does it work? Well – try it for a term that you know something about and see if it is erroneous. A quick lookup on "iPod," Apple's digital music player, yielded an entry that was informative and quite deep; less fulsome but still quite good was the entry for the "Peasants' Revolt" (including a reference to Barrie Dobson's excellent scholarly book).

The Wiki as Collaboration Tool

While Wikis that are password-protected strain the purist's definition of the form, they are commonplace as collaboration tools that allow a defined group (a project team, for example) to create, edit, and update documents collaboratively. The DLF makes use of

them in this manner in several working groups. There are some real benefits to the wiki as a collaborative tool, and some obvious limitations: one needs to learn the wiki text markup language, for example, in order to create layout instructions (emphasis, lists, tables, links, etc.), and there is nothing to stop multiple people editing the same document at the same time, with the last one to save the page being the one that survives (although the other edits are retained in the "history of changes" list).

Further Reading

Andrea Ciffolilli. "Phantom Authority, Self-Selective Recruitment, and Retention of Members in Virtual Communities: The Case of Wikipedia." *First Monday*, Vol. 8(12), December 2003. http://www.firstmonday.dk/issues/issue8_12/ciffolilli/

Ward Cunningham. "The Wiki." Microsoft Research – Social Computing Symposium 2004. 3/29/2004. A Web video presentation from Microsoft Research. [http://murl.microsoft.com/LectureDetails.asp?1076].

Brian Lamb. "Wide Open Spaces: Wikis, Ready or Not." *EDUCAUSE Review*, vol. 39, no. 5 (September/October 2004): 36-48. http://www.educause.edu/pub/er/erm04/erm0452.asp

Dave Mattison. "Quickiwiki, Swiki, Twiki, Zwiki and the Plone Wars. Wiki as a PIM and Collaborative Content Tool." *Searcher*, vol 11(4), April 2003: [http://www.infotoday.com/searcher/apr03/mattison.shtml]

WEB4LIB**:** August 2004: The Web4Lib discussion forum has had a series of online discussions about the value that wikis (and the Wikipedia) have to libraries. The archive of these discussions (ordered here by subject) can be found at: http://sunsite.berkeley.edu/Web4Lib/archive/0408/subject.html#start

David Weinberger. "Commentary: Wikis." *All Things Considered*, July 21, 2003. http://www.npr.org/features/feature.php?wfId=1344426 (text page leading to an audio file).

Appendix I: **Building and Distributing Resources in Atlantic World Studies.**
A Proposal for Collaboration between the DLF and Virtual Jamestown.

Summary

The Virtual Jamestown project is interested in pursuing a partnership with the Digital Library Federation. Under a planning grant from the Andrew W. Mellon Foundation, the project has created a database of online resources in Atlantic World Studies. Some of these collections are still being developed, such as the John Carter Brown Library Archive of Early American Images, many are already online. Some are online text databases, like the National Library of Canada, and others allow searches for maps of early Virginia, such as the Beinecke Rare Book and Manuscript Library at Yale University. Still others, such as the University of Kansas, AMDOCS: Documents for the Study of American History, serve as a portal to significant digitized documents: Columbus's journals; the Constitution of the Iroquois Nations; Richard Hakluyt, Ralph Lane, Thomas Hariot and John White from the Roanoke colony; links to early Virginia legislation; and other British colonial documents.

The Virtual Jamestown census of online collections found about 110 online sites with resources on Atlantic World Studies. We created a search engine to permit searches by keyword, place, people, site name, and by any word entered into the annotation category. The Atlantic World online resources database already includes at least four member institutions of DLF. In addition, some non-member institutions, such as the John Carter Brown Library, Huntington Library, and the British Museum with whom Virtual Jamestown has collaborated might also be brought into the DLF consortium. It just makes sense for Virtual Jamestown to build upon this database to include Atlantic World resources at DLF institutions and to include the capability of harvesting all the resources for those related to Jamestown.

A partnership with DLF would be a great asset in creating an Atlantic Studies Digital Archive (ASDA). Such an archive would have enormous impact on scholarship because it would permit scholars, especially in fields of Spanish, Dutch, Portuguese, English, and French history, access to a documentary, cartographical, archaeological, and visual/graphical database of immense value to global and comparative history. The ASDA would emancipate Jamestown studies from the narrow Virginia history focus to which it has been held captive and move it into the transatlantic world context where it was born, allowing scholars to assay Jamestown's significance and place in American and world history.

Common Pursuits

I see three areas where we have common pursuits:
- collections building
- tools development
- shared access to collections

Collections Building

The combined resources of DLF member institutions, Virtual Jamestown, and the Virginia Center for Digital History would collaborate in the building of an Atlantic Studies Digital Archive, pooling the vaults of content from member institutions of DLF and the existing Jamestown database of online resources into a large databank. The archive might then serve as a laboratory for developing common tools to harvest online collections and distribute the content according to guidelines and procedures mutually agreeable to all parties involved. Instead of just a few online collections now available through the Virtual Jamestown database, the DLF partnership could pull together under one umbrella the major corpus of material on Atlantic Studies from archives around the world.

Tools Development

DLF could take the lead in asking content providers to conduct a census of materials important to Atlantic Studies to be made available that are already digitized or those that institutions need assistance in bringing online. DLF could be the lead institution in brokering exchanges of existing datasets. The Virginia Center for Digital History, DLF, and specialists on the Virtual Jamestown project would work together to assemble the archive and experiment with tools development and infrastructure work. DLF could also broker the assembly of "grabbable" content into a federated tool that might exploit it or make it behave in ways researchers and teachers require for their work. Tools might be developed in collaboration with the providers who have already experimented with strategies to exploit their own collections with in-house tools that have potential for broader application, such as for example the "Scholar's Box" at the University of California, Berkeley. Another example is the CHART database that the Virginia Center for Digital History has created to facilitate the work of researchers interested in digitizing their material for the classroom or electronic publication. Scholars and teachers need a toolbox for exploiting content with features commonly recognized as essential but rarely integrated into a seamless ware for data manipulation and exploitation. The Atlantic Studies Digital Archive toolkit would provide the tools for such tasks as:

- searching
- note-taking
- creating databases for texts
- image manipulation
- map generation
- text annotation

Funders are far more likely to support tools development generously if they understand that the benefactors include all the major research libraries in the nation, instead of just those in Jamestown studies. A partnership in tools development would provide technical support to scores of scholars otherwise limited to the meager digital resources of their own institutions.

Enhanced Access

Once the Atlantic Studies Digital Archive is complete and tools for its use have been developed, all academic libraries who are content providers would share the distributed content with faculty and students at their respective institutions. There is nothing comparable to this kind of archive in the humanities. Imagine being able to send students to documents, maps, or images on population diasporas, indigenous-settler relations, the international slave trade, disease epidemics, or trade and consumerism where comparisons can be drawn, for example, from Africa, Canada, or South America under French, Dutch, Spanish, or Portuguese hegemony. Virtual Jamestown aspires to such universality and comparability, but could never achieve it on the scale that would be possible via a DLF partnership.

Contributions

Both Virtual Jamestown and DLF have developed funding sources, licensing agreements, and advisory boards of scholars and archivists. Again, it would make sense to combine these efforts and build upon the experience in grant writing, licensing, and governance.

Funding

Virtual Jamestown is already committed to a funding proposal that will include as key features support for collections building, technical assistance, and fellowship support on Jamestown and the Atlantic World. Potential funders include the Andrew W. Mellon Foundation, National Science Foundation, and the Institute of Museum and Library Services. As part of the proposal, Virtual Jamestown has donor funds of $50,000 to leverage its proposal to collaborate with teams of scholars in building electronic archives.

Licensing Agreement and Governance

A management consultant for Virtual Jamestown has made substantial progress in reaching consensus on issues of intellectual property and governance. In addition, the DLF has a model licensing agreement that could be extended to all institutions under this collaboration. An advisory board of prominent national and international scholars, digital archivists, and a management consultant meets four times a year to advise the project director, Dr Crandall Shifflett, professor of history at Virginia Tech, and Dr. Will Thomas, chairman of the board, director of the Virginia Center of Digital History and Associate Professor of History at the University of Virginia. We will meet in October to discuss a draft funding proposal and a business plan to maintain the Jamestown archive in perpetuity.

I hope we can talk soon about common goals and how we can combine our joint interests in collections building, tool development, and shared access to digital resources in Atlantic World studies.