# Table of Contents

2  RBS Course Syllabus **Andrew Stauffer and Bethany Nowviskie**

3  Sepulchres **Michael Furlough**

8  What We Talk About When We Talk About Repositories **Michael Furlough**

14  Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment **Constance Malpas**

34  XML, Interoperability and the Social Construction of Markup Languages: The Library Example **Jerome McDonough**

49  On Creating a Usable Future **Jerome McGann**

63  New-Model Scholarly Communication: Road Map for Change **Scholarly Communication Institute**

93  The Research Library in the 21st Century: Collecting, Preserving, and Making Accessible Resources for Scholarship **Abby Smith**

101  Storyboarding Examples and Templates **Adaptive Path**
Classes will be held in Alderman Library 317, with occasional trips to other venues. Please bring your laptop computer with you unless otherwise noted.

BRIEF SYLLABUS

MONDAY:
8:30am  Introductions and project/case study introductions.
10:30am  Librarians & humanists: communities of practice  
          (Read: SCI report, McGann)
1:30pm  Book work: bibliographic variety & “hard cases”
3:30pm  Site visit & discussion: Digital Curation Services

TUESDAY:
8:30am  Opportunities & challenges: the future of libraries  
         (Read: Malpas, Smith)
10:30am  Book work: evidence of use
1:30pm  Markup languages & networks (Read: McDonough)
3:30pm  Intro to digital design: user requirements/user experience

WEDNESDAY:
8:30am  Book work: editions & states.
10:30am  Comparing editions: facsimile, critical, digital
1:30pm  NINES & linked open data for the GLAM sector (guest speakers)
3:30pm  Hands-on design: storyboarding

THURSDAY:
8:30am  Book work: bindings & technologies of the book
10:30am  Non-textual data: the case of maps (guest speakers)
1:30pm  Neatline: geo-temporal interpretation of archival collections
3:30pm  Hands-on design: storyboard revisions

FRIDAY:
8:30am  Your storyboard presentations
10:30am  Sustainability & preservation: agents and stakeholders  
          (Read: Furlough)
2:00pm  Revisiting our first exercise: what are the issues now?
3:15pm  Retrospect & prospect
3:40pm  Course evaluation
4:00pm  Final reception, in the Press Room
I hate the word “repository” because it obscures the variety of problems we are attempting to address through their development, and in turn constrains our thinking about what may possible through the services they can enable. Modifiers such as “institutional,” “central,” “digital,” “open,” “collections” repository (or some torturous combination of these) do not help because each of these variations imply a singular technological solution to a set of complex changes in the way research is conducted and information is communicated. The term “repository” carries with it many connotations, some of them rather unfortunate. In general it describes a place where things lay, not where things are happening. According to the Oxford English Dictionary, a repository could be a “A vessel, receptacle, chamber, etc., in which things are or may be placed, deposited, or stored” (1.a). Definition 5—“A person to whom some matter is entrusted or confided”—is a less common usage, but one that certainly resonates with the institutional mission and responsibilities that libraries hold for their collections. Yet it is also hard to overlook definition 2.b: “A place in which a dead body is deposited; a vault or sepulchre.” [1]

The early energy surrounding institutional repositories (IR) centered on a hope that promoting open access could serve as a countermeasure to commercial publishing power and its ability to distort the market for knowledge. “Taking control” of our institution’s research by providing the ability to distribute this information to the world in an open access mode seemed to be an inevitable outcome of the Internet. Here is a brief history of institutional repository hype. In July 2002, The Chronicle of Higher Education reported “Superarchives’ Could Hold All Scholarly Output: Online collections by institutions may challenge the role of journal publishers.”[2] Also in 2002, a SPARC position paper declared

Institutional repositories—digital collections capturing and preserving the intellectual output of a single or multi-university community....[p]rovide a critical component in reforming the system of scholarly communication—a component that expands access to research, reasserts control over scholarship by the academy, increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to the institutions and libraries that support them.” [3]
But in 2004 The Chronicle provided an update: “Papers Wanted: Online archives run by universities struggle to attract material.” [4] IRs soon became the butt of jokes, even inside the community of practitioners. In March 2006, Dorothea Salo, an institutional repository manager, rechristened herself in her blog. “I have a new title. Innkeeper of the Roach Motel,” she wrote, describing her repository as a site where data goes in but doesn’t come out. [5] By November 2008, attendees at the SPARC Repositories Conference worried openly about how faculty can be persuaded to use the institutional repositories on their campuses and how these services were going to survive the worst economic crisis in decades if they didn’t.

Many of my publishing colleagues have warned me that if institutional repositories are successful, they will go out of business and eventually the entire scholarly communication system will start to break down. I can assure my friends that their jobs are quite safe. While IRs have generally had limited success, many publishers have adapted their policies to allow authors to distribute pre- or post- print versions of articles in open access forms. Those changes are at least partly related to funder and public pressure and the availability of repository outlets. Some institutions have begun to have luck negotiating with publishers for the rights to deposit their faculty’s articles in those same repositories. The emphasis on opening access has been driven heavily by institutional (library) hopes, not the needs of our users, whose work is changing, and who require new services to keep pace in their fields. Archiving single articles didn’t make much sense to them in that context. Continuing to focus on IR “deposit” by faculty and students—which sounds like a one-way proposition for the information—will not carry us forward. I am also not very hopeful about local, campus level “mandates” for open access, like those coming out of Harvard, MIT, and even the University of Kansas. It is hard work to establish a campus wide policy that defaults all researcher publishing to “open access,” and it’s easy to fail. Such mandates are great PR, but are they really enforceable at a local level? Are they really worth the time it takes to evangelize, combat falsehoods, smooth feathers and win converts? Will they really change the way that scholarship is communicated? In the end they are right thing to do, but they don’t really challenge anyone’s scholarly norms—in fact they go out of their way not to do so in order to win the political battle.

Repository tools and many related programs have been developed with a potential scope of use broader than that implied by the institutional repository hype, and may yet serve, as Clifford Lynch and Joan Lippincott wrote, as “general-purpose infrastructure within the context of changing scholarly practice.” [6] Deployment has varied. Some libraries have focused first on “the intellectual output of the institution,” while others have focused on particular disciplines or user groups, while still others have attempted to better manage and provide access to digitized versions of the physical collection of the library. Libraries are also using these services to manage “born-digital” resources acquired by the library from a variety of sources, including vendors and publishers. None of these activities are mutually exclusive, and it is likely that libraries will end up working with all of these materials simultaneously.
So what is it that we think we are talking about when we talk about repositories in research libraries today? Are repositories things? If so, they are more like conglomerate rock than uniform applications and programs. Are they places, like the open stacks or the closed archives? If so, they are Victorian follies—an aggregation of features, not all of them fully functional, offering none of the transparency of Phillip Johnson’s glass house. In the widest possible sense, when we are talking about repositories, we are talking about a set of organized methods for content management, not about specific applications or even specific access points online. Managing and providing access to diverse digital content requires many different processes, methodologies, policies and technologies, just like a physical library collection. Collectively, we are today determining how to manage digital data as smoothly and with the same degree of certainty as we do physical collections. [7] Repository-based content management can and must serve many functions at once, and successful implementations will recognize this to move beyond our early narrow focus to succeed. So where do we begin? One potential answer to these questions is provided by Catherine Mitchell, who at the 2008 SPARC conference presented with the title “Let’s stop talking about repositories,” arguing instead for a talk about services. [8] That is a small, but critically important rhetorical shift.

After all the hype, today it is most critical to identify the content-driven services that can be offered through “repositories,” and which of these our libraries need to offer to our clients, however we define them. I don’t believe that all libraries should offer such services. We are well past the days when all collections needed to reside physically on each campus, and we approaching times when replication of similar technology services on each campus may prove to be economically impossible. If content management and delivery services have a limited audience on a given campus, it may be better to partner with others to offer or to rent the needed technology. That is heresy to many, because it contradicts our philosophy of retaining control over “our” materials. But scale matters, and if we cannot it achieve it on our own we will risk poorly managing services that have limited use.

No library should implement a digital repository program without examining the role it will play in its broader strategy for collection development, stewardship, and providing access to its primary constituencies. That strategy should be based on a clear understanding of the community’s needs, and the requirements for long-term stewardship of the data collected. Most importantly, it should include a critical assessment of the library’s ability to fully meet those needs, including funding, the skills of its staff, and the benefit of the service relative to the cost of operating it. [9] We cannot do everything, especially now, and we should be willing to walk away from that which doesn’t work for us. As an administrator, I appreciate that this is much more easily said than done, and I have witnessed and been complicit in many situations where it has been necessary and expedient for political and “reputational” reasons to continue walking into the big muddy. We learn from failures, but institutions have a terrible time admitting them. Here’s to hoping that one effect of the Great Recession will be a greater willingness to walk away.

We tend to build silos for our collections and services, either because of
organizational politics, convenience, feasibility, or just because we are predisposed to think about fitting things into buckets. Some libraries that are offering significant services for original publications such as journals, for articles such as pre- or post-prints, and for large collections of reformatted or born-digital materials, operate each some or all of these services through different software and different operational divisions of that library. Heterogeneous content and heterogeneous communities require heterogeneous services, but a coherent organizational strategy and economies of scale should underlie these.

When we talk about repositories, or better, the services we offer through them, we are discussing the sociological side of technology and its adoption. Repository programs are still exotic, or even scary, to too many of our librarian colleagues, and to make things worse, most librarians were never trained to make the sale for experimental services or projects. But those programs must be integrated into the rest of the library’s services. Public services librarians meet students every day in the classroom, in the library, or online, and, in spite of their slight reluctance to paying us a visit in the library, faculty still call upon us. All of us have a responsibility to query our teaching and research colleagues to divine the needs that they didn’t know they had, and try to match those to the services we can provide. Asking questions that you don’t have answers to is the best way to start learning. That, in turn, requires more active communication across the divisions of our libraries to ensure that the programs we offer are integrated into instruction, reference and collection development.

Many researchers, perhaps scientists especially, cannot imagine why or how the library could do anything but subscribe to journals, even as they struggle to document and organize their work. We have huge obstacles to overcome, but the library remains a trusted brand, and our partners are out there and talking. Johanna Drucker wrote recently in the Chronicle of Higher Education:

> The design of new [online] environments for performing scholarly work cannot be left to the technical staff and to library professionals. The library is a crucial partner in planning and envisioning the future of preserving, using, even creating scholarly resources. So are the technology professionals. But in an analogy with building construction, they are the architects and the contractors. The creation of archives, analytic tools, and statistical analyses of aggregate data in the humanities (and in some other scholarly fields) requires the combined expertise of technical, professional, and scholarly personnel. [10]

In other words, we have to engage and guide researchers, but we must also let them lead us, possibly where we might not have expected, or maybe even wouldn’t want them to go. We can’t assume we know best, or the library will end up running a repository, i.e., “a place in which a dead body is deposited; a vault or sepulchre.”
NOTES


[7] Throughout this essay, I use the term “data” broadly to refer to just about anything that is in digital form and of enduring interest to scholars or librarians.


[9] Dorothea Salo has quite effectively written about the failure of institutional repository programs, attributing much of it to a failure of vision and leadership that results in a poor alignment of resources with the program goals. See Dorothea Salo, “Innkeeper at the Roach Motel “ Library Trends 57, no. 2 (2008).

What We Talk About When We Talk About Repositories

Mike Furlough, Guest Columnist

Correspondence concerning this column should be addressed to Diane Zabel, Schreyer Business Library, The Pennsylvania State University, 309 Paterno Library, University Park, PA 16802; e-mail: dxz2@psu.edu.

Mike Furlough is Assistant Dean for Scholarly Communications and Co-Director, Office of Digital Scholarly Publishing, The Pennsylvania State University, University Park, Pennsylvania.

In this column, Mike Furlough writes about repositories from a user services perspective. His engaging and accessible article provides a fascinating history of repository hype, a primer on technical tools, and thoughtful reflections on the future of institutional repositories. Mike Furlough joined The Pennsylvania State University Libraries in 2006 as the assistant dean for Scholarly Communications and co-director of the Office of Digital Scholarly Publishing. Furlough's graduate training is in American Literature, but he ran away to join the University of Virginia Library, where he developed and led a number of services to support digital scholarship. He currently serves as a member of the Association of College and Research Libraries' Scholarly Communications Committee and begins editing a column on that topic for C&RL News in 2009.—Editor

Throughout the past few years, I have come to dislike the word “repository” because it obscures the variety of problems we are attempting to address through their development, and in turn may constrain our thinking about what may be possible through the services they can enable. Modifiers such as “institutional,” “central,” “digital,” “open,” and “collections” (or some torturous combination of these) do not help because each variation implies a singular technological solution to a set of complex changes in the way research is conducted and information is communicated. “Repository” carries with it many connotations, some of them rather unfortunate. In general it describes a place where things lay, not where things are happening. According to the Oxford English Dictionary, a repository could be “a vessel, receptacle, chamber, etc., in which things are or may be placed, deposited, or stored” (definition 1.a). Definition 5—“A person to whom some matter is entrusted or confided”—is a less common use, but one that certainly resonates with the institutional mission and responsibilities that libraries hold for their collections. Yet it is also hard to overlook definition 2.b: “A place in which a dead body is deposited; a vault or sepulchre.”

“Institutional repository” (IR) often refers to a service that supports and encourages the deposit of student- and faculty-created materials, primarily open-access versions of research articles that have been formally published elsewhere or not at all. The early energy surrounding IRs centered on a hope that promoting open access could serve as a countermeasure to commercial publishing power and its ability to distort the market for knowledge. Taking control of our institutions’ research by providing the ability to distribute this information to the world in an open-access mode seemed to be an
What We Talk About When We Talk About Repositories

Inevitable outcome of the Internet. What follows is a brief history of IR hype.


Institutional repositories—digital collections capturing and preserving the intellectual output of a single or multi-university community—provide a critical component in reforming the system of scholarly communication—a component that expands access to research, reasserts control over scholarship by the academy, increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to the institutions and libraries that support them.3

But in 2004 The Chronicle provided an update: “Papers Wanted: Online Archives Run by Universities Struggle to Attract Material.”4 IRs soon became the butt of jokes, even inside the community of practitioners. In March 2006, Dorothea Salo, an institutional repository manager, rechristened herself in her blog. “I have a new title. Innkeeper of the Roach Motel,” she wrote, describing her repository as a site where data goes in but doesn’t come out.5 By November 2008, attendees at the SPARC Repositories Conference worried openly about how faculty can be persuaded to use the IRs on their campuses and how these services were going to survive the worst economic crisis in decades if they didn’t.

Many of my publishing colleagues have warned me that if IRs are successful they will go out of business, and eventually the entire scholarly communication system will start to break down. I can assure my friends that their jobs are quite safe. The emphasis on opening access has been driven heavily by our institutional (library) hopes, not the needs of our users, whose work is changing and who require new services to keep pace in their fields. Archiving single articles didn’t make much sense to them in that context. While IRs have generally had limited success, many publishers have adapted their policies to allow authors to distribute pre- or postprint versions of articles in open-access forms. Those changes are at least partly related to funder and public pressure and the availability of repository outlets. Some institutions have begun to have luck negotiating with publishers for the rights to deposit their faculty’s articles in those same repositories. However, continuing to focus on IR “deposits” by faculty and students—which sounds like a one-way proposition for the information—will not carry us forward.

Repository tools and many related programs have been developed with a potential scope of use broader than that implied by the IR hype, and may yet serve, as Clifford Lynch and Joan Lippincott wrote, as “general-purpose infrastructure within the context of changing scholarly practice.”6 Deployment has varied. Some libraries have focused first on “the intellectual output of the institution,” others have focused on particular disciplines or user groups, while still others have attempted to better manage and provide access to digitized versions of the physical collection of the library. Libraries also are using these services to manage born-digital resources acquired by the library from a variety of sources, including vendors and publishers. None of these activities are mutually exclusive, and it is likely that libraries will end up working with all of these materials simultaneously.

So what is it that we think we are talking about when we talk about repositories in research libraries today? Are repositories things? If so, they are more like conglomerate rocks than uniform applications and programs. Are they places, like the open stacks or the closed archives? If so, they are Victorian follies—an aggregation of features, not all of them fully functional, offering none of the transparency of Phillip Johnson’s glass house. In the widest possible sense, when we are talking about repositories, we are talking about a set of organized methods for content management, not about specific applications or even specific access points online. Managing and providing access to diverse digital content requires many different processes, methods, policies, and technologies, just like a physical library collection. Collectively, we are today determining how to manage digital data as smoothly and with the same degree of certainty as we do physical collections.7 Repository-based content management can and must serve many functions at once, and successful implementations will recognize this to move beyond our early narrow focus to succeed. So, where do we begin? One potential answer to these questions is provided by Catherine Mitchell, who at the 2008 SPARC conference presented with the title “Let’s Stop Talking About Repositories,” arguing instead for a talk about services.8 That is a critically important, rhetorical shift.

Repositories, Tools, and the Lifecycle of Digital Information

Both DSpace and Fedora, two major open-source repository tools developed over the past decade, attempt to cover many of the needs for effective data management and access in a repository service. Though these are sometimes referred to as repositories in themselves, other platforms have been deployed that are based on commercial, community-based, and homegrown applications. Each of these has limitations and requires their own tradeoffs of convenience and functionality for users and for system managers.

DSpace, first released in 2001 through a development partnership between Hewlett Packard and MIT, has been promoted since its inception as an application meant to develop an institutional collection of research. The earliest instance of the dspace.org website found in the Wayback Machine, dated April 28, 2001, greets the reader: “Welcome to DSpace, a newly developed digital archive created to capture and distribute the intellectual output of MIT.”9 DSpace provides a set of integrated tools, services, and functions designed to make repository start-up simple, which has led to broad adoption: over the last decade DSpace has been put into production at
FOR YOUR ENRICHMENT

more than five hundred institutions around the world.

Fedora is an acronym for Flexible Extensible Digital Object Repository Architecture, which sums up the project’s philosophy of enabling maximum flexibility and adaptability in the design, implementation, and use of its software. Fedora was initially conceived by researchers at Cornell University’s Computer Science Department and later developed in partnership with the University of Virginia Library and with funding from the Andrew W. Mellon Foundation. From the start, Fedora has been marketed as the foundation for a wider variety of digital collection management needs, not as an integrated IR solution.

Both DSpace and Fedora recognize the needs of diverse disciplines and researchers. DSpace’s interfaces define collections relevant to communities with different needs and expectations for distributing digital content online. Administrators have the ability to enable variable controls on input and access and support different formats, genres, and metadata structures to describe and document those materials. Fedora’s flexibility allows each instance to serve unique purposes designed for the case at hand; no two installations look alike or serve the same purpose. Until recently, these two tools have developed and been managed independently. However, the DSpace Foundation and Fedora Commons announced on May 12, 2009, that they would merge and form a new organization called DuraSpace. DuraSpace will continue to support and develop DSpace and Fedora and also develop new services to work with both platforms. More details about this change can be found at http://duraspace.org/pressrelease.html.

These are not the only two repository tools in use. Some libraries offer IR services using commercially developed software. The University of Utah has deployed an IR service using ContentDM, a product offered by OCLC and originally created to help organizations manage digital library collections of images and other reformatted materials. Digital Commons, a product of BePress, is a hosted solution for IR programs, providing libraries with an opportunity to offer programs with limited technology investment. The California Digital Library was an early adopter of Digital Commons and uses it for their eScholarship platform for all of the University of California campuses. Organizations with unique missions and more resources may develop comprehensive archival systems through a variety of applications and technologies, many of them specifically designed and tailored for the particular mission. The National Archives and Records Administration is now developing the Electronic Records Archive, a comprehensive electronic archives management system to handle and preserve the electronic records of the Federal Government. The HathiTrust, a shared repository service, is developing its own infrastructure for managing the collective digital content of more than twenty different libraries, projected to include five hundred thousand newly digitized volumes each month.

These repository tools and implementations of them are, or should be, part of larger systems and strategies for building and caring for collections. Some models of archival and scholarly practices can help to elucidate this. The Open Archival Information Systems (OAIS) Reference Model, an ISO standard, provides a high-level conceptual overview of the organizational and technological functions necessary for the effective archival management of digital data. The OAIS model identifies four core activities in a repository system: ingest (methods to define, describe, document, and authorize the transfer of digital files); data management (the capture, storage, and analysis of metadata); archival storage (infrastructure to protect the integrity of the files at the byte level); access (provided to the user through queries, retrieval, and viewing, or to other applications or archival systems). Governing these activities are preservation planning (to develop strategies to mitigate risk and monitor technological change) and administration (such as the policy decisions that define the goals of the archive or collections and provide the support through financial and other resources). The OAIS framework describes not software or tools, but principles of practice that should support their use.

Another model more explicitly addresses our interactions with the data we collect and manage. The Digital Curation Centre at the University of Edinburgh has developed a Curation Lifecycle Model, which vividly captures what we do with information resources in libraries and in our colleges and universities. The lifecycle is a continuous flow of activity where information, represented in digital objects, is selected for acquisition, made accessible, discovered, used, transformed, reacquired and distributed, discovered anew, continuously appraised, and sometimes disposed of. This model reminds me of the “rip-mix-burn” credo associated with the Free Culture movement, which can also be crudely applied to the processes of research and scholarship: a remixing of information to create new knowledge. Our faculty and students inquire, discover, and sort information resources then analyze and synthesize them into new work, which is written, published, and distributed for the next scholars. Curation is an active process, one in which our users can and should participate. For a simple example, consider that Fedora has been designed with the assumption that digital objects have multifaceted and overlapping relationships between themselves, and that identifying and making these relationships explicit is a part of scholarship and archival work. A digital object may thus belong to many different networks of content, rather than one parent grouping, and would be accessible through all of those organizing contexts in the repository. A library might deposit a set of images as a defined collection (perhaps all coming from a single source or supplier). Users may wish to re-present those same images as constituent parts of many other collections or sets (such as “images of Italy” or “images for the Western Art Survey”) and to ensure that those relationships and representations are defined, recorded, persistent, and discoverable for other users.

Libraries are moving from a business model based primarily on managing the products and the output of research and scholarship to a model based on facilitating the process of
What We Talk About When We Talk About Repositories

scholarship, teaching, and research that result in those products. No matter what technology is used, the lifecycle model for digital data curation suggests that repositories and digital data management are not distinct backroom technology operations but activities that should be functionally integrated into the mission and services of the library. Repository-enabled services will be critical to the future of scholarship in general, regardless of who offers them. Commercial agents, such as Google, can outperform existing library systems on speed and breadth of basic searches, but the preservation and scholarly use of digital assets are still fertile ground for libraries, technologists, and library users.

WHAT DO WE MEAN BY SERVICES?

For many of us, the answer to this question is obvious: Services are the activities we perform to support the researchers, students, teachers, and members of the public who use our libraries. But the term “service” also has a specific use, referring to technical functions conducted through interoperable, machine-to-machine interfaces. For instance, many digital repository collections use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to automatically share their metadata with other collections with the aim of improving discovery. Such services have little direct human intervention and provide additional functionality and data to applications, including repository systems.

Some IR programs may have developed as services looking for a need, but nothing creates a need for their services like an institutional or legal mandate to use them. Since the spring of 2008, several elite universities or their colleges— including Harvard’s Faculty of Arts and Science and John F. Kennedy School of Government, Stanford’s School of Education, and all of MIT—have adopted policies requiring that all publications of their faculty also be made available in an open-access online service managed by the library. When the National Institutes of Health (NIH) began to require PubMed deposit of all publications based on NIH funding, several universities took advantage of the change to offer enhanced digital data curation suggests that repositories and digital data management are not distinct backroom technology operations but activities that should be functionally integrated into the mission and services of the library. Repository-enabled services will be critical to the future of scholarship in general, regardless of who offers them. Commercial agents, such as Google, can outperform existing library systems on speed and breadth of basic searches, but the preservation and scholarly use of digital assets are still fertile ground for libraries, technologists, and library users.

WHAT DO WE MEAN BY SERVICES?

For many of us, the answer to this question is obvious: Services are the activities we perform to support the researchers, students, teachers, and members of the public who use our libraries. But the term “service” also has a specific use, referring to technical functions conducted through interoperable, machine-to-machine interfaces. For instance, many digital repository collections use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to automatically share their metadata with other collections with the aim of improving discovery. Such services have little direct human intervention and provide additional functionality and data to applications, including repository systems.

Some IR programs may have developed as services looking for a need, but nothing creates a need for their services like an institutional or legal mandate to use them. Since the spring of 2008, several elite universities or their colleges—including Harvard’s Faculty of Arts and Science and John F. Kennedy School of Government, Stanford’s School of Education, and all of MIT—have adopted policies requiring that all publications of their faculty also be made available in an open-access online service managed by the library. When the National Institutes of Health (NIH) began to require PubMed deposit of all publications based on NIH funding, several universities took advantage of the change to offer enhanced services to their community. For example, the University of Michigan performs NIH deposit on behalf of their own researchers, using their existing IR deposit methods as a way of gathering the articles for submission to PubMed Central. Other mandates will prove more complicated to meet, such as archival electronic records management. In this arena, a variety of public policy and regulatory matters have driven technology companies such as Sun, Hewlett-Packard, and EMC to develop new storage technologies that can better ensure the integrity of digital data making up the records of businesses, financial institutions, and government. Higher education institutions must meet similar requirements for auditing and disclosure, and this emerging need will challenge organizations traditionally tasked with maintaining the records and archives. The diversity of formats, record types, and relationships between these records creates a challenging environment in which to establish basic policies and practices to ensure compliance with legal requirements.

Mandates are not a reliable growth model for services. However, aggregating content is our bread-and-butter, and repository systems can enable large-scale aggregation to offer improved access. At the University of Virginia Library, the Fedora repository platform was concurrently developed with an integrated digital collections system that uses Fedora to deliver electronic texts, images, and special collections finding aids. The metaphor here would be the stacks, rather than the archive. Among the texts are books digitized by the Library alongside full-text databases published by ProQuest that are also included in their Literature OnLine (LION) product. The image collections include both purchased and licensed sets from Archivision, as well as images digitized from Virginia’s Special Collections and contributed by faculty at the school. Bringing these disparate collections into an integrated collection management environment enables searches across the collections and makes it possible to create additional applications that allow users greater functionality. The Collectus tool, for example, provides a way to save sets of images and texts for use and sharing with their classes or colleagues.

Some institutional repository services and their infrastructure serve as the basis for publication activities. Campus-based publishing has become an increasingly visible (though still very experimental) service at many research libraries and smaller ones as well. They share core assumptions with broader IR programs: Libraries, working with faculty and often with publishers such as university presses, can provide cost-effective technology to support the open distribution of research literature from within the university. Compared with IRs, these programs require an even greater degree of faculty engagement while offering a more specific service focused on distributing complete titles or collections. In her 2008 study for the Association of Research Libraries (ARL), Research Library Publishing Services, Karla Hahn reported that 44 percent of eighty responding ARL libraries offered some form of publishing service for journals, monographs, or conference proceedings. Open-source publishing tools such as Open Journal Systems and DPubS are frequently used for these purposes and sometimes are used to provide support for editors and authors in their review and submission processes. A large number of institutions reported using DSpace or Digital Commons for their publishing platform. Although DSpace does not offer native, out-of-the-box workflow tools for editors of publications, Digital Commons offers users editorial workflow tools designed for BePress’s own journals. Given the experimental nature of these efforts, it appears that many institutions are limiting costs by first taking advantage of their existing technology investments before investigating more specialized service offerings.

Publishing, in the limited sense of distribution, can be integrated into the curriculum via repository programs as well. Like many others, Penn State—my own university—now requires that all theses and dissertations be submitted to the Graduate School in electronic format through a system managed by the libraries. One simple model to extend this...
FOR YOUR ENRICHMENT

could include providing an electronic deposit service to under-graduate programs that require a formal thesis or paper for graduation. The Ethnography of the University Initiative (EUI) at the University of Illinois Urbana–Champaign offers another model. Through classes associated with EUI, undergraduates in different fields of study engage in original research about their campus using their familiar home environment to explore the concepts they are learning. EUI provides these students with experience publicly distributing their work by selecting research reports for inclusion in IDEALS, the University Library’s repository service. The EUI collection, numbering more than 350 works, also serves as a research collection for students engaged in the program. 

Less formal, direct-to-reader publishing via media such as blogs has become an increasingly important part of daily discourse and scholarly communication. Blog software is readily available, and some institutions have created a centralized service specifically for their students and faculty. The ephemeral nature of most blogging is reflected in the software: these are first and foremost authoring and distribution tools, and do not provide the all of the capabilities for data management, preservation, and discovery that we expect to see in an archival service. The National Science Digital Library (NSDL), an initiative funded by the National Science Foundation, has launched a blog service known as Expert Voices, which features postings by multiple experts to promote online collaboration on science topics across different communities, such as K–12, researchers, and librarians. NSDL has designed Expert Voices so that it can easily interoperate with other resources in the library and so that the discussions and new content may be directly captured and managed in a repository environment based on Fedora. Rather than sitting off to the side, the data created in the blogs can easily become a part of the managed digital library.

Even in areas of traditional, formalized publication, computationally intensive research has begun to affect the nature of the research article, its relationship to the journal as an outlet, and its relationship to the primary evidence that supports its argument. Figures, tables, and other graphics have provided a means to adumbrate evidence in the context of a scholarly argument in print. In the sciences, new observational tools and high performance computing resources have made possible the capture and analysis of research data on a very large scale that often extends beyond a single university or nation. Such research, and its related outputs, poses tremendous challenges for universities, libraries, and scientists to devise stable environments for secure, long-term access. At Johns Hopkins University, the Libraries have worked closely with astronomers based at Hopkins and with the National Virtual Observatory to explore what technological capacity and organizational expertise will be required to capture and manage astronomical data. An element of this project has included work with major publishers in the field to define and test persistent methods of publishing experimental data sets along with the formal written article reporting on the research. Here the library has begun to develop a service that supports not only their local community but potentially the entire profession of astronomy.

CONCLUSIONS

After all the hype, today it is most critical to identify the content-driven services that can be offered through “repositories,” and which of these our libraries need to offer to our clients, however we define them. I am of two minds about whether all libraries should offer such services. On the one hand, we do have a mission to collect, manage, preserve, and provide access to resources for our user base and the wider world. But we are well past the days when all collections needed to reside physically on each campus, and we are approaching times when replication of similar technology services on each campus may prove to be economically impossible. If content management and delivery services have a limited audience on a given campus, it may be better to partner with others to offer or to rent the needed technology. That is heresy to many because it contradicts our philosophy of retaining control over “our” materials. But scale matters, and if we cannot achieve it on our own, we will risk poorly managing services that have limited use.

No library should implement a digital repository program without examining the role it will play in its broader strategy for collection development, stewardship, and providing access to its primary constituencies. That strategy should be based on a clear understanding of the community’s needs and the requirements for long-term stewardship of the data collected. Most importantly, it should include a critical assessment of the library’s ability to fully meet those needs, including funding, the skills of its staff, and the benefit of the service relative to the cost of operating it. We cannot do everything, especially now, and we should be willing to walk away from that which doesn’t work for us. As an administrator, I appreciate that this is much more easily said than done.

We tend to build silos for our collections and services, either because of organizational politics, convenience, feasibility, or just because we are predisposed to think about fitting things into buckets. Some libraries that are offering significant services for original publications—such as journals, for articles such as pre– or postprints, and for large collections of reformatted or born-digital materials—operate some or all of these services through different software and different operational divisions of that library. Heterogeneous content and heterogeneous communities require heterogeneous services, but a coherent organizational strategy and economies of scale should underlie these.

Access services to repository content present their own challenges. The relative quality of the user interfaces and overall user experience of interacting with our tools and systems continues to be a problem throughout all library systems. California Digital Library has begun an effort to revamp user interfaces and functionalities in their eScholarship repository—which is unusual given their use of a commercial service provider, BePress—but this takes many resources to
do effectively. I once solicited feedback from a faculty member using a cross-collection metadata search for a digital library collection. Comparing it to “those finding aids they make me use in special collections,” he said it looked like librarians created it for librarians. Ouch.

Then there is the issue of authenticity and value, and what that means in different contexts and for different communities. While software can help to ensure that the digital file’s integrity remains stable, it is still difficult to identify and explain variations in the multiple versions of materials deposited in different repositories (Google Scholar can at least identify related versions, however). Some repository services don’t easily support versioning, and we might be reluctant to withdraw items deposited, even if an author offered a substantial revision in its place. But if we deliver a draft of an article in our repository, but perhaps don’t subscribe to the outlet producing the finished product, how do we help our users know how to evaluate what they find in our collections? Here public services, collection development policies, and technology programs need to work in concert to help convey the context for what is found online in our electronic collections.

When we talk about repositories, or better, the services we offer through them, we should be discussing the social side of technology and its adoption. Repository programs are still exotic, or even scary, to too many of our colleagues, and most librarians were never trained to make the sale for experimental services or projects. But those programs must be integrated into the rest of the library’s services. Public services librarians meet students every day in the classroom, in the library, or online, and, despite their slight reluctance to pay us a visit in the library, faculty still call upon us. All of us have a responsibility to gently query our teaching and research colleagues to divine the needs that they didn’t know they had, and try to match those to the services we can provide. That, in turn, requires more active communication across the divisions of our libraries to ensure that the programs we offer are integrated into instruction, reference, and collection development. In some fields, perhaps especially in the sciences, many researchers cannot imagine why or how the library could do anything but subscribe to journals, even as they struggle to document and organize their work. We have huge obstacles to overcome, but the library remains a trusted brand and our partners are out there and talking. Johanna Drucker, a professor of information studies at the University of California at Los Angeles, wrote recently in the *Chronicle of Higher Education* that

> the design of new [online] environments for performing scholarly work cannot be left to the technical staff and to library professionals. The library is a crucial partner in planning and envisioning the future of preserving, using, even creating scholarly resources. So are the technology professionals. But in an analogy with building construction, they are the architects and the contractors. The creation of archives, analytic tools, and statistical analyses of aggregate data in the humanities (and in some other scholarly fields) requires the combined expertise of technical, professional, and scholarly personnel.\(^2\)

In other words, we have to engage and guide researchers, but we also must let them lead us, possibly where we might not have expected, or maybe even wouldn’t want them to go. We can’t assume we know best, or the library will end up running a repository, i.e., “a place in which a dead body is deposited; a vault or sepulchre.”

### References and Notes

7. Throughout this essay, I use the term data broadly to refer to just about anything that is in digital form and of enduring interest to scholars or librarians.
13. Though it considered contracting with the Harris Corporation to build a system based on Fedora, it ultimately awarded a contract to Lockheed Martin, which is designing the system on that basis of a variety of internally developed methods and applications. See www.archives.gov/era (accessed July 29, 2009).
Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment

Constance Malpas

Program Officer
OCLC Research
Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment
Constance Malpas, for OCLC Research

© 2011 OCLC Online Computer Library Center, Inc.
Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA): http://creativecommons.org/licenses/by-nc-sa/3.0/.
January 2011

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

OCLC (WorldCat): 695086590

Please direct correspondence to:
Constance Malpas
Program Officer
constance_malpas@oclc.org

Suggested citation:
Contents

Acknowledgments ........................................................................................................................................ 7

Executive Summary .................................................................................................................................... 8

Introduction ............................................................................................................................................... 13
  Premise .................................................................................................................................................. 14
  Methodology ......................................................................................................................................... 14
  Scope of Analysis .................................................................................................................................. 15

Summary of Findings .................................................................................................................................. 17
  Shared Digital Repository Profile: HathiTrust .................................................................................... 17
  Shared Print Repository Profile: ReCAP .............................................................................................. 32
  Model Consumer Profile: NYU ............................................................................................................. 45

Shared Print Provision: Assessing the Options ....................................................................................... 50
  Expanding the Scope of Shared Service .............................................................................................. 50
  Assessing Market Maturity .................................................................................................................. 51
  Alternative Service Providers ............................................................................................................... 52
  Optimizing Existing Infrastructure ...................................................................................................... 55

What is It Worth? Putting a Price on Shared Collection Services ......................................................... 58

Conclusions and Recommendations ...................................................................................................... 64

Appendix I. HathiTrust Cost Rationale .................................................................................................. 67
Appendix II. Cloud Library Service Agreements: ReCAP as Shared Print Repository ..................... 71
References ........................................................................................................................................76
Figures

Figure 1. Growth of HathiTrust Digital Library collection (June 2009 - June 2010) ........... 18

Figure 2. Projected growth of HathiTrust Digital Library (June 2010 - June 2020) ........... 19

Figure 3. Primary document types of titles in HathiTrust Digital Library (June 2010) ........ 20

Figure 4. Distribution of HathiTrust Digital Library titles by document type (June 2009 - June 2010) .......................................................................................................................... 21

Figure 5. Subject distribution of titles in HathiTrust Digital Library (June 2010) ............ 22

Figure 6. Distribution of titles in HathiTrust Digital Library by subject and copyright status (June 2010) .................................................................................................................. 27

Figure 7. Top ten categories of public domain content in HathiTrust Digital Library (June 2010) ................................................................................................................................. 29

Figure 8. System-wide distribution of library holdings for titles in HathiTrust Digital Library (June 2010) ...................................................................................................................... 31

Figure 9. Distribution of ReCAP holdings by contributor (July 2010) ............................. 33

Figure 10. Growth in titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010) ........................................................................................................ 34

Figure 11. Primary document types of titles duplicated in ReCAP and HathiTrust Digital Library (June 2010) ...................................................................................................................... 37

Figure 12. Subject distribution of Hathi titles held in ReCAP (June 2010) ...................... 38

Figure 13. Comparative scope of shared digital and shared print repository collections (June 2010) ................................................................................................................................. 40

Figure 14. Titles duplicated in ReCAP and the HathiTrust Digital Library (June 2010) .... 42

Figure 15. System-wide distribution of library holdings for Hathi titles in ReCAP (June 2010) 44
Figure 16. Growth in coverage of NYU Bobst holdings in HathiTrust Digital Library (June 2009 - June 2010) ................................................................. 46

Figure 17. NYU Bobst titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010) ................................................................. 47

Figure 18. NYU Bobst titles duplicated in UC SRLF and HathiTrust Digital Library (June 2009 - June 2010) ................................................................. 53

Figure 19. Comparison of potential shared print provision options for NYU Bobst Library (June 2010) ................................................................. 54

Figure 20. NYU Bobst titles duplicated in ReCAP partner libraries and HathiTrust Digital Library (June 2009 - June 2010) ................................................................. 56

Figure 21. Percentage duplication of titles held in ARL libraries and HathiTrust Digital Library (June 2009 and June 2010) ................................................................. 62
Into being
The clouds condense, when in this upper space
Of the high heaven have gathered suddenly,
As round they flew, unnumbered particles—
World’s rougher ones, which can, though interlinked
With scanty couplings, yet be fastened firm,
The one on other caught.

Lucretius *De rerum natura*, Book V
trans. William Ellery Leonard (1921)

**Acknowledgments**

The Cloud Library project emerged out of a series of discussions that began with Carol Mandel, Jim Neal, John Wilkin and Jim Michalko in 2009. These individuals provided leadership and vision that guided all the work that followed.

Library staff from New York University, Columbia University, the New York Public Library and Princeton University participated in a variety of meetings, conference calls and e-mail exchanges that helped to give shape to the project. The Andrew W. Mellon Foundation contributed financial support under a grant ably administered by Chuck Henry at the Council on Library and Information Resources (CLIR).

Michael Stoller, Bob Wolven, Zack Lane, Matthew Sheehy, Marvin Bielawski and Eileen Henthorne made essential contributions to the project, not least in helping to compile ReCAP holdings data for inclusion in our analysis. Kat Hagedorn and Jeremy York provided expert technical and operational support from Hathi. Jenny Toves ensured that WorldCat data extractions were available on schedule.

I am grateful to Jim Michalko, John Wilkin and Paul Courant for their many thoughtful questions and suggestions about the data analysis and interpretation. Lorcan Dempsey and Brian Lavoie also provided insights and helpful methodological guidance along the way.

Particular thanks are due to Roy Tennant and Bruce Washburn, who provided expert programming support over the course of this project and routinely produced small miracles, and to Patrick Confer for his diligent editorial work in preparing the final report.
Executive Summary

The Cloud Library project was jointly designed and executed by OCLC Research, the HathiTrust, New York University’s Elmer Holmes Bobst Library, and the Research Collections Access & Preservation (ReCAP) consortium, with support from The Andrew W. Mellon Foundation. The objective of the project was to examine the feasibility of outsourcing management of low-use print books held in academic libraries to shared service providers, including large-scale print and digital repositories.

The following overarching hypothesis provided a framework for our investigation:

- The emergence of a mass-digitized book corpus has the potential to transform the academic library enterprise, enabling an optimization of legacy print collections that will substantially increase the efficiency of library operations and facilitate a redirection of library resources in support of a renovated library service portfolio.

From this, a number of research questions emerged:

- What is the scope of the mass-digitized book corpus in the HathiTrust Digital Libray and to what degree does it replicate print collections held in academic research libraries?

- Can public domain content in the HathiTrust Digital Library provide a suitable surrogate for low-use print collections in academic libraries?

- Is there sufficient duplication between shared print storage repositories and the HathiTrust Digital Library to permit a significant number of academic libraries to optimize and reduce total spending on local print management operations?

- What operational gains might be obtained through a selective externalization of collection management activities?

Based on a year-long study of data from the HathiTrust, ReCAP, and WorldCat, we concluded that our central hypothesis was successfully confirmed: there is sufficient material in the
mass-digitized library collection managed by the HathiTrust to duplicate a sizeable (and growing) portion of virtually any academic library in the United States, and there is adequate duplication between the shared digital repository and large-scale print storage facilities to enable a great number of academic libraries to reconsider their local print management operations. Significantly, we also found that the combination of a relatively small number of potential shared print providers, including the Library of Congress, was sufficient to achieve more than 70% coverage of the digitized book collection, suggesting that shared service may not require a very large network of providers.

Analysis of the distribution of subject matter and library holdings represented in the HathiTrust Digital Library and shared print repositories further confirmed that the digital corpus is largely representative of the collective academic library collection, suggesting a broad potential market for service. A further positive finding was that monographic titles in the humanities constitute the greatest part of the mass-digitized resource, which may indicate that some relatively under-resourced disciplines will begin to benefit from a digital transformation that has already powered enormous innovation in the sciences. As detailed below, we also found that substantial library space savings and cost avoidance could be achieved if academic institutions outsourced management of redundant low-use inventory to shared service providers.

Our findings also revealed some important obstacles and limitations to implementing changed print management practices in the current library operating environment. The following are among the most important constraints we identified:

- The proportion of public domain content in the HathiTrust Digital Library is relatively small (approximately 16% of titles in June 2010) and typically represents material that is not widely held in the library system; as a result, the number of libraries that might hope to reduce local print management costs for these titles through negotiated agreements with the HathiTrust and shared print providers is quite low. Moreover, the age and subject distribution of titles in the public domain is not representative of academic research collections as a whole. In sum, the public domain corpus as currently defined by U.S. copyright law cannot be considered a viable surrogate for any academic print collection.

- While significant duplication was found between the HathiTrust Digital Library and multiple large-scale library storage collections, it was apparent that no single print storage repository could offer coverage sufficient to enable significant space savings or cost avoidance for a given client library. Put another way, effective shared print storage solutions will depend upon a network of providers who will need to optimize holdings as a collective resource.
The absence of a robust discovery and delivery service based on collective print storage holdings is an impediment to changed print management strategies, especially for digitized titles in copyright.

It is our strong conviction, based on the above findings, that academic libraries in the United States (and elsewhere) should mobilize the resources and leadership necessary to implement a bridge strategy that will maximize the return on years of investment in library print collections while acknowledging the rapid shift toward online provisioning and consumption of information. Even, and perhaps especially, in advance of any legal outcome on the Google Book Search settlement, academic libraries have a unique opportunity to reconfigure print supply chains to ensure continued library relevance in the print supply chain. In the absence of a licensing option, online access to most of the digitized retrospective literature will be severely constrained. Demand for print versions of digitized books will continue to exist and libraries will be motivated to meet it, but they will need to do so in more cost-effective ways. In the absence of fully available online editions, full-text indexing of digitized in-copyright material provides a means of moderating and tuning demand for print versions and should facilitate the transfer of an increasing part of the print inventory to high-density warehouses. Viewed in this light, shared print storage repositories could enable a significant and positive shift in library resources toward a more distinctive and institutionally relevant service portfolio.

Our study assessed the opportunity for library space saving and cost avoidance through the systematic and intentional outsourcing of local management operations for digitized books to shared service providers and progressive downsizing of local print collections in favor of negotiated access to the digitized corpus and regionally consolidated print inventory. As detailed in the report that follows, the organizational change required to achieve these gains is likely to be substantial and challenging to implement. Yet, the opportunity costs of inaction may prove even greater than the risks of enacting shared print management regimes. Many of the positive transformations that academic library directors hope to achieve in the next decade or so will require a fundamental shift in collections management. The scope and scale of change that is possible may be judged by these key findings:

- As of June 2010, the median rate of duplication between titles held by university libraries in the U.S. Association of Research Libraries (ARL) and the HathiTrust Digital Library exceeds 30%; that is to say, nearly a third of the content purchased by research-intensive libraries in the United States has already been digitized and is preserved in a shared digital repository.

- If the current growth trajectory of the HathiTrust Digital Library is sustained, we can project that more than 60% of the retrospective print collections held in ARL libraries
will be duplicated in the shared digital repository by June 2014. This growth rate far exceeds average annual acquisitions in ARL libraries, suggesting that the digital replication of legacy collections will outpace growth of new physical collections, enabling a transformation in traditional library operations, staffing and space requirements.

- The median space savings that could be achieved at an ARL library if a robust shared print offer were in place today amounts to approximately 36,000 linear feet or the equivalent of more than 45,000 assignable square feet (ASF). These are conservative estimates based on the assumption that holding libraries own a single copy of each duplicated title. Actual space savings could be much greater. In practical terms, this means each library could recover space sufficient for a learning or research commons, media lab, or office space for faculty and visiting scholars.

- The total annual cost avoidance that could be achieved if shared print service provision for mass-digitized books were available today would amount to a figure between $500,000 and $2 million per ARL library, depending on the physical environment (e.g., open stacks on campus or high-density off-site storage) in which the titles would be managed locally.

Academic library directors can have a positive and profound impact on the future of academic print collections by adopting and implementing a deliberate strategy to build and sustain regional print service centers that can meet aggregate demand with aggregate supply. Beyond the obvious operational efficiencies of consolidating low-use, digitized print volumes into shared service collections there is an important strategic advantage to reconfiguring collective inventory that is increasingly devalued as an institutional asset. A proactive effort to rationalize collections that are undergoing a radical phase change from print to digital will enable libraries to achieve a careful and measured wind-down of operations that no longer deliver distinctive value, while continuing to uphold a vital preservation and access mandate.

The shared infrastructure needed to support a broad-based externalization of legacy print management functions is unlikely emerge without directed action and decision-making by leaders in the academic library community. Individuals and organizations interested in advancing these changes are encouraged to consider the following recommendations:

*Library directors and managers can . . .*

- Advocate in favor of licensed access to the mass-digitized resource as part of a comprehensive strategic plan in which the library can reassert its role as a vital part of the academic enterprise.
• Engage directly with faculty and academic officers to communicate a compelling strategy in which selective externalization of traditional functions is demonstrably improving the institution’s ability to fulfill an academic and research mission.

• Support the HathiTrust’s ongoing efforts to expand public access to the mass-digitized book corpus by affiliating with the organization as a content contributor or sustaining partner.

Prospective shared print providers, including managers of large print storage facilities, can . . .

• Proactively build collections that will deliver maximum operational value to external audiences; leverage the collective library investment in mass digitization and the HathiTrust by accelerating the transfer of mass-digitized titles to print preservation repositories.

• Contribute to the establishment of a common service profile by surfacing model agreements and engaging in community dialog about the operational and business requirements of shared service provision.

Research organizations, including OCLC Research, Ithaka S+R, JISC and other similar entities, can . . .

• Advance our collective understanding of the changing profile of demand for legacy print collections in the mass-digitized environment.

• Help to characterize the optimal redistribution of library resources in different regional and national contexts.

Funding bodies, including IMLS, the Mellon Foundation, NEH and others, can . . .

• Provide funding to support the implementation of shared print management through grants to libraries and other organizations to subsidize the direct costs of title selection and processing until such activities are fully subsumed as ongoing library operations.
Introduction

In spring 2009, a group of ARL directors came together to discuss a common set of challenges and opportunities facing university libraries and identify some shared strategies for responding to them. A number of circumstances were converging that appeared to offer some potential relief from critical space pressures in the library and the increasingly burdensome operations associated with managing a large local inventory of low-use print collections.

The seemingly imminent resolution of the Google Book Search settlement was an important motivating factor: academic libraries were confronting the prospect, at once daunting and liberating, of licensed access to a massive aggregation of digitized books from major U.S. research collections. Would such a collection substantially duplicate local print holdings? If so, what consequences might ensue for traditional academic library operations?

At the same time, the emergence of the HathiTrust, a shared digital repository consolidating much of the library-contributed content from the Google Books database, appeared to resolve many of the concerns the library community had regarding long-term stewardship of the mass-digitized book corpus. In combination with the large aggregations of low-use print collections managed in high-density library storage facilities, Hathi might bridge the gap between a well-documented decline in the use of academic print collections and the anticipated shift toward scholarly reliance on full-text electronic resources.

The fact that critical elements of the shared infrastructure needed to effect a large-scale transition from print to electronic research collections were owned and managed by the library community itself gave library directors confidence that the timing and outcomes of this transition could be managed according to the needs of the academic community and not dictated by the business objectives of commercial providers. Were the combined resources of Hathi and large-scale shared print providers already sufficient to mobilize a change in library operations? What was the scope of service likely to be? How much and what kind of value would it need to deliver? Who—which kinds of libraries and in what number—would benefit? These questions were compelling enough to justify a joint research project in which potential service providers and consumers could explore business requirements, service expectations and feasibility of implementation.
The initiative that emerged from these discussions within ARL came to be known as the “Cloud Library” project, because it posited a future in which library collections and services would be sourced from external providers, reducing local infrastructure and operational expenditures in a manner analogous to the cloud-sourced business and computing solutions that now prevail in the commercial and high-tech sectors. Funded by The Andrew W. Mellon Foundation, the project was staffed by a team of investigators from the HathiTrust, the Research Collections Access and Preservation consortium (ReCAP), New York University Libraries, and OCLC Research. This report provides a high-level summary of findings from this project.

**Premise**

The research questions that motivated this study reflect a conviction shared by all of the participating institutions: the emergence of a mass-digitized book corpus has the potential to transform the academic library enterprise, enabling an optimization of legacy print collections that will substantially increase the efficiency of library operations and facilitate a redirection of library resources in support of a renovated library service portfolio. We started from the presumption that academic libraries will be motivated to transfer resources (space, personnel, and capital) from local print management operations to shared print and digital repositories in proportion to the tangible benefit that cooperative management confers. We were therefore less interested in examining the theoretical advantages of shared service provision than in characterizing the operational gains (space recovery and cost avoidance) that might be obtained through a selective externalization of collection management activities.

**Methodology**

Between June 2009 and June 2010, a monthly snapshot of records was harvested by OCLC Research from the publicly available HathiTrust metadata repository. These records were machine-processed to extract OCLC numbers and, where necessary, to extract and map alternative identifiers (LCCN, ISBN or ISSN) to valid OCLC numbers. The resulting batch of OCLC numbers was used to extract bibliographic records and holdings data from the WorldCat database each month. These bibliographic master records were then merged with selected Hathi metadata and (starting in September 2009) a sample of associated ReCAP repository customer codes to produce a single, consolidated dataset for analysis.

A master database was built to support analysis of the compiled data, which was programmatically enhanced to support analysis of key attributes of the aggregate collection, including broad subject areas, total library holdings, institutional source of the digitized text and copyright status. This database was enriched each month with successive snapshots of the
Hathi repository, mapped to WorldCat holdings and ReCAP customer codes as described above. By June 2010, the project database comprised 37 million records, representing a longitudinal view of the growing corpus of library-owned titles that are duplicated in print and digital repositories.

**Scope of Analysis**

In the twelve months covered by this project, the HathiTrust Digital Library doubled in size, increasing from approximately 3 million volumes to more than 6 million volumes. *On a per-volume basis, the shared digital repository is now larger than the average ARL library collection*; the median reported holdings at university-based ARL libraries in 2008 was approximately 3.5 million volumes. Because our analysis of the HathiTrust collection focuses on unique titles (manifestations or editions), rather than physical items, the number of records we compiled each month was somewhat smaller than the number of records in the Hathi metadata repository. Not every volume in the HathiTrust represents an individual book or journal title, and there is at least some duplication in content ingested from different contributors; as a result, the total number of volumes in the Hathi repository is more than the number of titles covered in our analysis. In June 2009, we identified approximately 2 million unique titles in the HathiTrust Digital Library; by June 2010, that number had grown to more than 3.6 million titles. For purposes of comparison, this represents *a collection comparable in scope to research libraries in the top tier of the U.S. ARL rankings*, based on holdings set in the WorldCat database. Indeed, at the time of writing, the number of unique titles in the HathiTrust Digital Library exceeds the number of titles cataloged and held by many research libraries.

A key goal of this research project was to assess the scope of coverage in shared print and shared digital repositories, with a view to understanding how the combined resources might enable a local reduction in redundant print inventory. For this reason, it was important to understand how much of the print storage collection in ReCAP is duplicated—or is likely to be duplicated—in the HathiTrust Digital Library. As of this writing, the shared ReCAP facility holds more than 8 million items contributed by the three partner libraries. Since the ReCAP collection is not currently visible as a discrete set of holdings in WorldCat, and building a union catalog of ReCAP holdings was beyond the scope of this project, we based our analysis on a representative sample of ReCAP holdings supplied by Columbia University and NYPL. Taken collectively, Columbia and NYPL’s ReCAP holdings amount to more than 75% of current inventory and this was deemed to be sufficient for our analysis.

The sample supplied to us included a broad range of materials managed under 14 different ReCAP customer codes, each representing a different set of request and circulation rules. The large size and broad scope of the sample gave us reasonable confidence that findings from our
analysis could be generalized across the ReCAP collection as a whole. Storage, selection and transfer protocols at the three partner libraries are based on common parameters (low use monographs; journals duplicated in electronic format), so that the nature, if not the content, of the materials contributed by each is likely to be comparable.

To provide a baseline against which duplication of ReCAP holdings in the HathiTrust Digital Library might be assessed, we periodically compared patterns in the ReCAP sample against other large-scale print storage collections that are more readily subject to analysis in WorldCat. Findings from these analyses are presented below.
Conclusions and Recommendations

Our year-long study of the mass-digitized book corpus in the HathiTrust Digital Library and parallel investigation of potential shared print service providers has confirmed that there is an opportunity for significant library space savings and cost avoidance if management operations for digitized books are deliberately and systematically outsourced or externalized to shared service providers.

One can anticipate that academic institutions interested in reducing local print holdings in favor of regionally consolidated inventory will, in years to come, increasingly look to extant repositories like ReCAP, the UC Regional Library Facilities, etc. as a source of preservation and access services. Our findings suggest that current storage inventory is not presently optimized to support shared print solutions on a large scale, but also indicates that system-wide reorganization of collections and services that maximizes the business value of print as a cooperative resource is both feasible and capable of producing great benefit to the academic library community.

Our findings suggest that the shared infrastructure that is needed to support a broad-based externalization of legacy print management functions is unlikely to materialize without some purposeful action by the academic library community. By describing and—where possible—quantifying the value that a changed infrastructure might deliver, we hope to have contributed in some measure to stimulating potential consumers and suppliers alike. Further work will be needed before academic libraries and the educational institutions they serve can fully realize the benefits of shared service:

- It is in the interest of all academic libraries that mass-digitized collections be made more widely available to students and researchers, and that their scope and quality improve to the degree that low-use print inventory can be retired in favor increased reliance on digital surrogate. Library directors and academic administrators should advocate in favor of licensed access to the mass-digitized resource as part of a comprehensive strategic plan in which the library can reassert its role as a vital part of the academic enterprise.
The HathiTrust’s ongoing efforts to expand public access to the mass-digitized book corpus through programmatic rights assessment, direct negotiation with rights holders, and by accessioning large aggregations of digitized public domain resources, should be recognized as a major contribution to the transformation of the library service environment. By developing a partnership model that is not dependent on content contribution, Hathi can deliver benefit to an even broader range of academic institutions.

- **Beginning in 2013, the HathiTrust will introduce a new cost model that will enable non-content contributing members to participate in (and benefit from) joint stewardship of the digital repository. This new affiliation model is described in Appendix I.**

- Institutions and organizations that aspire to roles as shared print service providers need to proactively build collections that will deliver maximum operational value to external audiences; they should leverage the collective library investment in Hathi by accelerating the transfer of mass-digitized titles to print preservation repositories and self-consciously promote these resources as shared and cooperatively managed assets. Above all, these institutions must take steps to make their service ambitions and capacity known, so that potential consumers can begin to articulate core requirements and common service profiles can be identified.

  - **In cooperation with other library organizations, OCLC is working to develop technical solutions that will enable the latent value of library storage collections and distributed print archives to be more effectively disclosed.**

- Prospective shared print providers can help to define a common service profile by surfacing model agreements and engaging in dialog about the operational and business requirements of shared service provision. Managers of large print repository collections should be empowered and encouraged to engage in business modeling exercises that are explicitly intended to expand the market for service beyond content contributors.

  - **ReCAP partner libraries have outlined core elements of a shared service agreement. These are summarized in Appendix II.**

- Libraries that are already motivated to outsource legacy print management functions in support of a changed service portfolio or simply to relieve local space pressures should begin to establish objective targets, and quantify and articulate desired outcomes so that motivated suppliers can respond in kind. Library administrators should engage directly with faculty and academic officers to communicate a compelling strategy in which selective externalization of traditional functions is
demonstrably improving the institution’s ability to fulfill an academic and research mission. This work will be challenging and deserves external support and endorsement by library leadership organizations and funders.

- Research organizations can advance our collective understanding of the changing profile of demand for legacy print in the mass-digitized environment and help to characterize the optimal redistribution of library resources.

When these steps are taken, we will have made measurable progress toward the worthy goal of ensuring the long-term survivability of the scholarly record at a cost that is sustainable for the research library community as a whole.
References


\[\text{References}\]


XML, Interoperability and the Social Construction of Markup Languages: The Library Example

Jerome McDonough <jmcdonou_at_illinois_dot_edu>, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign

Abstract

The past decade has seen XML widely adopted within a variety of communities, including the digital library community. While it now plays a critical role in the infrastructure of many digital library operations, XML's promise of interoperability of data across systems and organizations has not been fully realized within digital libraries. The reasons for this are not primarily technical in origin, but social, and relate to the cultures of XML's designers and XML language implementors, and a failure on the part of the digital library community to grapple with the sociotechnical nature of XML and its implementation. Possible strategies for addressing these issues of interoperability might include reduction of the flexibility afforded by specific XML-based markup languages used by the digital library community, and an increased focus on standardizing translations between various communities of practice use of such markup languages.

Introduction: Failures of Interoperability with XML

Eleven years after its endorsement by the World Wide Web Consortium, XML has been widely adopted within numerous, disparate communities and in a vast range of application domains, from standards for electronic filing of federal income tax [Internal Revenue Service 2007] to user interface design [Goodger et al. 2001]. The digital library community has been an active and early adopter of XML, for use in structuring both content and metadata. The reasons for this rapid uptake of XML within the digital library community are familiar to anyone with experience in the world of markup languages:

- XML helps ensure platform (and perhaps more critically vendor) independence, simplifying the migration of content between systems;
- XML provides the multilingual character support critical to the handling of library materials;
- XML's extensibility and modularity allow libraries to customize its application within their own operating environments;
- XML helps minimize software development costs by allowing libraries to leverage existing, open source development tools;
- XML, through virtue of being an open standard which enables descriptive markup, may assist in the long-term preservation of electronic materials; and perhaps most importantly
- XML provides a technological basis for interoperability of both content and metadata across library systems.

For all of these reasons, XML-based content standards such as the Text Encoding Initiative (TEI) have seen wide adoption within the library community, and librarians have been actively engaged in the development of a number of XML-based metadata standards, including Encoded Archival Description (EAD), Metadata Object Description Schema (MODS), Metadata Authority Description Schema (MADS), Metadata Encoding and Transmission Standard (METS), Metadata for Images in XML (MIX), MPEG-21 Digital Item Declaration Language (DIDL), Open Archives Initiative Object Reuse and Exchange (OAI-ORE), Preservation Metadata...
Implementation Strategies (PREMIS) and many others. XML is now used throughout the research library world, and is a fundamental part of the infrastructure developed within the digital library community over the past decade.

Despite its success, however, XML has not lived up to many librarians’ expectations within one area, that of interoperability. Efforts to exchange information employing XML-based metadata standards such as Dublin Core have fallen prey to a number of encoding and semantic inconsistencies [Shreeves et al. 2005]. Variations in the use of namespacing (date vs. dc:date), in regional conventions (08-11-2008 in the U.S. vs. 11-08-2008 in the U.K.) and language and culture (11 Août, 2008 vs. 8 Sha'aban 1429 A.H.) confound application developers trying to process XML data.

Perhaps more surprising is the failure of XML to ensure interoperability at a syntactic level. Digital library developers have expected that shared use of a XML standard for structuring content and metadata (what is commonly called "structural metadata" within the digital library community) would ensure content interoperability and promote the development of tools and services designed to work with content encoded according to that standard [Hurley et al. 1999]. In practice, however, this goal has proved rather elusive. Experiments conducted by participants in the Library of Congress National Digital Infrastructure for Preservation Program (NDIIPP) to test the exchange of digital objects between repositories failed even when participants were using the same XML-based encoding format [DiLauro et al. 2005], [Shirky 2005].

While some of the failures experienced by the Library of Congress NDIIPP tests were the result of incompatible repository infrastructure, others resulted from mismatched expectations regarding the appropriate use of METS, one of the XML formats employed for the test. DiLauro et al., discussing Johns Hopkins University's experience in the NDIIPP tests, state,

Stanford commented after their ingest of the JHU archive that they had expected one METS object for the entire archive. Because our approach resulted in many METS files – on the order of the number of items in the archive – the Stanford ingest programs experienced out-of-memory conditions. This situation may have been ameliorated had they used the reference code provided by JHU; however, this will be an area that we will look into for future archive ingest projects.

This matter points to a broader issue observed during the various import processes of this phase. Though three of the four non-LC participants (including JHU) used METS as part of their dissemination packages, each of our approaches was different. Clearly there would be some advantage to working toward at least some common elements for these processes. [DiLauro et al. 2005]

As alluded to by [DiLauro et al. 2005], a critical difficulty for achieving interoperability using structural metadata standards such as METS is the level of flexibility they enable in structuring a description of an object. As [Nelson et al. 2005] note in their discussion of using the MPEG-21 Digital Item Declaration Language during the NDIIPP test, it is possible to map a single object into multiple different encodings using MPEG-21, depending on the level of granularity you wish to employ in the description. The same is true of METS and other, similar information packaging standards. They each provide a grammar to describe the structure of complex digital objects. To facilitate description of arbitrarily complex structures, these standards employ a relatively flexible grammar, and document authors can and do find a variety of different ways to describe the structure of a single object using one of those grammars.

To date, the digital library community has treated these interoperability issues surrounding structural metadata standards as a technical problem demanding a technical solution. Most efforts to solve these interoperability problems have focused on the use of a profiling mechanism to further constrain the creation of instance documents conforming with a XML schema, sometimes in conjunction with a validation mechanism (such as Schematron) to test instance documents conformance with the additional requirements established in the profile [Littman 2006], [Keith 2005]. In essence, profiles exist to limit authors' flexibility in the use of a given XML language. If different institutions can agree on using a particular profile of a language, they are far more likely to be able to produce content objects which can be readily exchanged and interoperate with a variety of different systems.

However, while these mechanisms may be successful in ensuring interoperability within a narrowly defined local context, they are not in themselves any guarantee of interoperability at the scale envisioned by digital
library projects such as Aquifer [Kott et al. 2006], which hope to promote the ready exchange and interoperability of digital library content among a multitude of institutions. The official METS profile registry already contains a variety of mutually incompatible profiles for similar types of objects, with profiles varying in their choices of descriptive metadata (Dublin Core vs. MODS), use of controlled vocabularies in descriptive and administrative metadata sections, and their structure (e.g., requiring the use of a single structMap element in the case of the Oxford Digital Library METS Profile and mandating the use of two structMap elements in the case of the Indiana University METS Navigator profile). While profiles may enable localized interoperability, they do not necessarily lead to widespread agreement regarding the best ways of describing objects' structure, and in fact, it is conceivable that to a certain extent they reify differences between institutions. Allowing the specification of local profiles of a XML language may help formalize the problem of interoperability, but it does not solve it.

If we are to deal with the issues of interoperability that continually manifest themselves in the realm of structural metadata standards for digital libraries, we need to cease viewing this purely as a technical problem, and acknowledge that it is the result of the interplay of technical and social factors. The XML standards for structural metadata employed by the digital library community represent cases of sociotechnical systems, and only when we have analyzed the social, as well as the technical, components of these systems will we be able to suggest how they may be optimized to achieve the goals of interoperability, usability and preservability desired by librarians and their patrons.

**XML from a Sociotechnical Perspective**

One of the fundamental tenets of sociotechnical systems theory is that technological design and technological evolution are not value neutral processes. Technological design is both informed by, and seeks to inform, the social environment in which technology is used, and the work of designers and engineers can be seen as being as much social engineering as technical engineering. By providing a new means of accomplishing a task, a technologist is also prescribing a new set of behaviors centered on the new technology (and possibly proscribing others). This conceptualization of technology was concisely summarized by [Akrich 1992], who argues that

...when technologists define the characteristics of their objects, they necessarily make hypotheses about the entities that make up the world into which the object is to be inserted. Designers thus define actors with specific tastes, competences, motives, aspirations, political prejudices, and the rest, and they assume that morality, technology, science, and economy will evolve in particular ways. A large part of the work of innovators is "inscribing" [emph. original] this vision of (or prediction about) the work in the technical content of the new object [Akrich 1992, 207–8]

All the existing and developing standards for structural metadata within the digital library community are XML-based. Any sociotechnical examination of these standards therefore must start with at least some consideration of XML itself. Our questions concerning XML, then, are what world view have XML's authors inscribed within it and what influence has that had on XML's uptake and use within the digital library community.

We can learn a great deal about the viewpoints of a particular technology's designers from the documents they author which define the goals for the technology (e.g., use cases and user needs analysis), those which help implement the technology (e.g., specification documents), and those which attempt to explain or promote the new technology to potential users. If we look at the original specification document for XML, we find a relatively clear set of goals for the technology enumerated:

1. XML shall be straightforwardly usable over the Internet.
2. XML shall support a wide variety of applications.
3. XML shall be compatible with SGML.
4. It shall be easy to write programs which process XML documents.
5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero.
6. XML documents should be human-legible and reasonably clear.
7. The XML design should be prepared quickly.
8. The design of XML shall be formal and concise.
9. XML documents shall be easy to create.
These goals convey some of the world view that XML’s designers brought to bear in creating the technology. They saw XML as a transmission format for communications (hence the requirement that it be usable over the Internet). They believed that XML’s success was contingent upon it being flexible enough to "support a wide variety of applications." They also clearly believed that for XML to succeed it must be easy to use, but they also recognized that the meaning of "ease of use" is contingent upon the use one might make of the technology. Ease of use for a document author ("XML documents shall be easy to create," "XML documents should be human-legible and reasonably clear") is a good deal different from ease of use for a software developer ("It shall be easy to write programs which process XML Documents," "The design of XML shall be formal and concise").

This relatively small set of goals for the XML language was further elaborated upon by members of the original World Wide Web Consortium (W3C) XML Working Group in a variety of papers they authored to introduce and clarify XML to its potential user community. [Bosak 1998] further defines the goals of XML as supporting the user needs of "extensibility, to define new tags as needed," "structure, to model data to any level of complexity," "validation, to check data for structural correctness," "media independence, to publish content in multiple formats," and "vendor and platform independence, to process any conforming document using standard commercial software or even simple text tools." The benefits adhering to XML’s providing a standardized format are identified as including "complete interoperability of both content and style across applications and platforms; freedom of content creators from vendor control of production tools; freedom of users to choose their own views into content; easy construction of powerful tools for manipulating content on a large scale; a level playing field for independent software developers; and true international publishing across all media.”

Emancipatory language is invoked repeatedly here through the use of the terms freedom and independence, particularly affording users the freedom "to define new tags" and in so doing "choose their own views into content." In the designers’ world view, a key benefit to XML is the freedom it provides users to define their own structure for documents and data, using their own semantics, and to escape restrictions that software vendors (through their own inscriptions in their own products) might wish to impose on their users. Other articles by members of the XML Working Group (see, for example, [Bosak 1997] and [Bosak & Bray 1999]) reiterate a vision of XML as a technology allowing users to define their own structures while simultaneously supporting interoperability of documents and data on a global scale.

The XML 1.0 Recommendation bears the inscription of its designers’ ideological stance towards appropriate mechanisms for data and document structuring as well as appropriate relationships between document creators and software and platform vendors. The effort to promote a metalanguage over any specific markup language, the adoption of Unicode as a basic character set, and the elimination of SGML features which proved difficult to implement (including CONCUR, OMITTAG and SUBDOC) are some of the technological means through which XML’s designers sought to normalize and reify a particular set of social and technological relationships. Nor did this process stop with the release of the XML 1.0 recommendation in 1998. The period between February 1998 and October 2001 saw the development and release of a plethora of additional XML specifications, including XML Namespaces, XSLT, XPath, XML Schema, XLink/XBase, XML Information Set and XSL-FO, as well as a variety of XML software tools including parsers, editors and stylesheet engines. All of these various technological objects presented their own opportunities for their designers to further refine the ideological inscription carried within the XML 1.0 Recommendation. One of these objects in particular, the Namespaces in XML Recommendation, deserves further examination due to its significant affect on structural metadata standards developed by the digital library community.

[World Wide Web Consortium 1999a] provides the following justification for the introduction of a formal namespace mechanism into XML:

We envision applications of Extensible Markup Language (XML) where a single XML document may contain elements and attributes (here referred to as a "markup vocabulary") that are defined for and used by multiple software modules. One motivation for this is modularity; if such a markup vocabulary exists which is well understood and for which there is useful software available, it is better to re-use this markup rather than re-invent it.

Such documents, containing multiple markup vocabularies, pose problems of recognition and collision. Software modules need to be able to recognize the tags and attributes which
they are designed to process, even in the face of "collisions" occurring when markup intended for some other software package uses the same element type or attribute name. [World Wide Web Consortium, 1999b]

A strong motivating force for the "Namespaces in XML" recommendation, then, was a desire to promote modularity in the design of markup languages. Interoperability was also cited as a motivating factor by the World Wide Web Consortium in the introduction of "Namespaces in XML" [World Wide Web Consortium, 1999b]. Fundamentally, the authors of the "Namespaces in XML" recommendation wanted to simplify XML document authors' lives by ensuring that they did not need to reinvent markup languages which already existed, and that they could readily mix elements and attributes conforming to disparate schemas within a single document instance without worrying about collisions between element and attribute names. Again, modularity and flexibility in design of markup languages would give users the freedom they need while also insuring interoperability.

XML's designers have inscribed two overarching messages within the technology they have created. The first is that XML is about establishing a new social relationship between content creators and software vendors. By putting control of data formats into the hands of the content creation community via an open standard, XML provides that community with significant political leverage. They can avoid the proprietary data formats that software vendors have used to lock them into continuing use of a particular software package. XML thus represents the path to freedom. The second message is that XML enables easy communication and interoperability. XML will not only allow you to control your content, it will make it easier for others to use your content. Freedom and interoperability are the two underlying themes running through the complete set of XML specifications, with modular design embraced as the means for achieving these ends.

**Structural Metadata Standards and the Digital Library Community**

Libraries' exploration of the use of markup languages for encoding of library data predates the origin of XML by several years. Collaborations with the digital humanities community on the development of the TEI Guidelines, the development of Encoded Archival Description, and early efforts to apply SGML to bibliographic data[4] provided the library community with experience in the use of markup languages and demonstration of the benefits they could provide. When the XML 1.0 Recommendation was released, many digital library projects were already using SGML, and libraries were quick to embrace XML. XML's simpler design meant that software tools for processing XML data were readily available, and the new capabilities for data typing introduced by XML schema languages made XML even more attractive for certain uses than its predecessor, SGML. Early projects which employed XML, such as the Making of America II project [U.C. Berkeley Library 1997], were rapidly followed by a number of XML-based markup languages intended for use in the library community. A significant focus of much of the library community's work with XML has been developing languages which can serve to structure all the metadata and data comprising a digital library object into a coherent whole. Examples of languages developed and explored for this purpose in the library community include the Metadata Encoding and Transmission Standard (METS), the Fedora Object XML (FOXML) language, MPEG-21 Digital Item Declaration Language, and the new Open Archives Initiative Object Reuse and Exchange (OAI-ORE) specification.

Most of these languages employ a similar pattern for structuring content and metadata. They provide an encoding mechanism which allows the author to record a hierarchical structure defining the object, and then associate both content and metadata with various nodes within that structure. Figure 1 depicts a very simple version of such a structure, a book with a single chapter; metadata and content files (and metadata for the content files) are associated with appropriate nodes.

**Figure 1.**

A simple digital object

A METS encoding for such an object can be seen in Example 1. The hierarchical structure for the object is defined within the structMap element as a set of recursive div elements. Subsidiary fptr elements within a div are used to associate that div element with content files described in separate file elements, and ID/IDREF linking attributes are used to associate the root div element with a descriptive metadata record, and the individual file elements with an administrative metadata record. A TYPE attribute on the div elements allows the METS document author to indicate the type of subobject represented by each node in the structural hierarchy.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<mets xmlns="http://www.loc.gov/METS/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink"
/standards/mets/mets.xsd">
  <dmdSec ID="DM1">
    <mdWrap MDTYPE="OTHER">
      <xmlData>
        <meta>Descriptive Metadata for Book</meta>
      </xmlData>
    </mdWrap>
  </dmdSec>
  <amdSec>
    <techMD ID="AM1">
      <mdWrap MDTYPE="OTHER">
        <xmlData>
          <meta>Administrative metadata applicable to TIFF files</meta>
        </xmlData>
      </mdWrap>
    </techMD>
  </amdSec>
</mets>
```
The goal of representing the structure of a work as a hierarchy of nested div elements with TYPE attributes was to have a relatively simple, abstract hierarchical structure that could be readily applied to a variety of materials. This was intended to promote the adoption of the standard (a single, simple standard is more likely to be adopted than a variety of complex ones), which in turn was seen as promoting interoperability. Having all of the digital library community using a single standard for structuring content and metadata was seen by METS' designers as preferable to the community adopting a disparate set of standards.

It should be noted that this move towards abstraction was a relatively significant break from the SGML design practices that many research libraries had been using to date. While it is true that the notion of using nested div elements was derived from the TEI text apparatus, TEI does not rely on pure abstraction; one does not expect to encounter encoding such as div type="figure" in a TEI document, when a figure element is available to use. Just as XML itself, the METS schema carries an inscription of its designers' world view, that it was preferable to develop a single, simple, generalizable, highly abstract model and encoding mechanisms to structure content and metadata for digital library objects, rather than to pursue the development of a variety of highly specific schemas (one for photographs, one for journals, etc.), or a grand encompassing schema that contained elements appropriate to different genres that could be combined as needed (e.g., the TEI model). Despite its use of the abstract div element with a TYPE attribute to represent the structural components of a digital library object, however, the METS schema insisted on the use of more specific concrete elements to identify different forms of metadata, with the dmdSec element used for descriptive metadata and the amdSec element used for administrative metadata, along with a series of subelements for different forms of administrative metadata (technical, rights, provenance and source). This typification of different forms of metadata was itself an effort to promote both modularity in further metadata schema development and the creation of certain types of metadata schema. By identifying specific subclasses of metadata within the METS schema, METS' designers hoped to encourage XML developers in the digital library world to create discrete, specialized metadata standards that would align with those subclasses, and that those creating digital library objects could then select from a set of such modular XML metadata standards in composing a particular object. Through METS' design, its implementers consciously sought to encourage the adoption of modular schema design practices within the digital library community.

Other XML-based markup languages adopted by the digital library community have taken a similar approach. The MPEG-21 Digital Item Declaration Language also employs a rather abstract hierarchical structural mechanism for ordering content and metadata. It differs inasmuch as non-structural metadata (Descriptor elements in MPEG-21 parlance) are not typed, and structural metadata elements, while still rather abstract, are of three different types: Container, Item and Component. The Open Archives Initiative Object Reuse and Exchange specification is perhaps the most abstract of all the structural metadata standards adopted within the digital library community; while it has multiple serialization syntaxes, all of them employ a single mechanism to
link an abstract aggregation with a set of aggregated resources (although the specific linking mechanism varies according to the serialization syntax employed). Those aggregated resources may in turn be aggregations, and any aggregation may be associated with a variety of additional metadata.

If we examine these other standards to determine what inscriptions their designers have placed within them regarding their use, we find messages very similar to that of METS. Structural metadata should be highly abstract, so that a very small set of elements can be employed to structure widely disparate content genres. While METS was perhaps more vocal in trying to push the message that further development of metadata schemas should try to create small, focused and modular metadata sets that could be drawn upon as needed to encode a particular object, the other standards convey the same message (through the use Descriptor elements to associate metadata with other elements in the case of MPEG-21, and through RDF mechanisms in the case of OAI-ORE). Other structural metadata standards of interest to the digital library community employ similar mechanisms. The XFDU standard for data archiving uses hierarchies of ContentUnit elements that may be associated with dataObjects and metadataObjects. The IMS Content Packaging standard for learning objects uses hierarchies of item elements that may be associated with resources and metadata. While implementations differ in details, the pattern is similar and widespread across the various structural metadata standards of interest to the digital library community. Again and again we see designers seeking to achieve wide adoption of their standard in order to promote interoperability between differing institutions; to secure this goal, they favor a highly abstract structural mechanism which can be applied to a wide variety of content, and mechanisms to allow a variety of additional metadata schemas to "plug and play" within the larger structural framework.

While of perhaps some limited interest to social researchers of technology, none of the preceding seems particularly surprising or problematic. That the designers of XML itself, and the designers of encoding standards for digital library metadata and content, should favor flexibility, extensibility, modularity and the use of abstraction to support the generalizability of their standard, and hence promote its widespread adoption to help achieve interoperability, would not be a great shock to anyone who has spent more than five minutes in the company of computer scientists. These are all considered almost innate goods among software engineers in general and markup language enthusiasts in particular. Yet the NDIIPP tests cited previously would seem to indicate that flexibility, extensibility, modularity and abstraction are not in and of themselves sufficient to achieve interoperability. So what, specifically, is the problem that METS and other structural metadata standards are encountering?

**Defining the Problem, or, Why is XML like a rope?**

One of the earliest discussion points in the development of the METS standard was which of the various elements within the schema should be declared mandatory and which optional. After some discussion among the members of the working group that established METS' original design, it was decided that the structMap element, which records the basic tree structure on to which content files and metadata are mapped in METS, would be the only required element. METS, in the group's opinion, was fundamentally a structural metadata standard; it existed to provide a framework into which other descriptive and administrative metadata, as well as content, could be placed. The structMap element provided the tree upon which all the other structural components of METS where hung, where the logical and physical structure of a work could be delineated, and so was really the only section that needed to be mandatory. As the structMap was the only mandatory portion of a METS file, it was also expected that any structural description of a work should reside there; software that would process a METS file would expect to find logical or physical descriptions of the structure of a work residing within a structural map, and not elsewhere in the METS file.

It was a matter of some surprise for many in the METS community, then, when the Library of Congress, which serves as the maintenance agency for the METS standard, began to produce METS files for digital versions of certain kinds of audio recordings which placed the logical structure of the works in MODS records contained within the METS descriptive metadata section (dmdSec) rather than in the structural map, and registered a profile of METS establishing this as their formal internal practice for "recorded events" [Library of Congress 2006]. The MODS record within a METS file would provide a logical structure for the work using a hierarchical arrangement of the MODS relatedItem element, while the METS structMap would contain the physical structure, with ID/IDREF links used to draw connections between the two structural descriptions. A recorded concert, for example, might have a MODS record containing a hierarchy such as this. [8]
Meanwhile, the structural map would record the physical structure of the work as follows:

```
<div TYPE="cd:compactDiscObject" DMDID="MODS1">
  <div TYPE="cd:disc">
    <div DMDID="DMD_disc01_tr001" TYPE="cd:track"></div>
    <div DMDID="DMD_disc01_tr002" TYPE="cd:track"></div>
    <div DMDID="DMD_disc01_tr003" TYPE="cd:track"></div>
    <div DMDID="DMD_disc01_tr004" TYPE="cd:track"></div>
    <div DMDID="DMD_disc01_tr005" TYPE="cd:track"></div>
    <div DMDID="DMD_disc01_tr006" TYPE="cd:track"></div>
  </div>
</div>
```

This represents a valid use of METS (in the technical XML sense), but is a departure from expected practice, which would be to include both the logical and physical structural information within one or more structMap elements.

This example demonstrates two problems that have impeded the development of interoperable content within the digital library community. The first is that the implementation of highly abstract elements for the definition of structure provides a tremendous amount of flexibility to document encoders; there are a vast number of potential encodings of any given object in METS, with variations possible in depth of structure (do I limit my structure to musical movements or do I provide structural information down to the measure level?), labeling...
The second problem is what we might call the problem of standards independence. Many of the XML metadata schemas that have been developed with the help of the digital library community have been created with the understanding that ensuring their usefulness in a variety of application environments requires that they not contain inherent dependencies on other schemas; they need to be able to express all the relevant information within their particular domain on their own. In many of these XML standards, the designers recognized a need to be able to account for relationships between various content objects being described, whether the description being applied was the more traditional form of intellectual description you would expect in a library catalog, or a technical description of the composition of a TIFF image. The result has been that a number of common metadata schemas within the digital library field contain elements for expressing structural metadata, even schemas that are not primarily intended for recording structural metadata. Dublin Core has its relation element, MODS has its RelatedItem element, the PREMIS schema for preservation metadata has a relation element, even the MIX standard for still image technical metadata contains an element for referencing previous image metadata. As the standards’ developers felt they should not make their efforts dependent on structural metadata mechanisms in other standards, they implemented their own. Unfortunately, with the addition of each new metadata standard containing structural metadata capabilities, the potential for difficulties with our first problem increases. Every new metadata standard created within the digital library community seems to add another mechanism for describing the structural relationships between content objects, and hence greater potential for variation in object encoding practices.

The irony is that both these problems derive from the flexibility, extensibility, modularity and use of abstraction to create structural metadata elements that the designers of the metadata schemas hoped to promote. The potential range of variation in encoding structural metadata is the result of each of these factors. The use of abstraction in METS (i.e., the div element) was an attempt to make the standard flexible in application; however, it opens up a tremendous degree of play in encoding practice. If you ask two different individuals how many page breaks there are in a text, the likelihood that they will give the same answer is a good deal greater than if you ask them how many divisions there are in a text. The use of abstraction opens up encoding to a much greater degree of personal interpretation, and hence variation. The extensibility of METS, and the hope to promote a modular system of metadata schema reuse that its authors inscribed within it, opens up the possibility of using other metadata schemas to encode structural metadata. And it was this same desire for flexibility and modularity that has led other metadata schema designers to include structural metadata components in their own schemas; they wanted to ensure that their own efforts were flexible enough to be applied in a variety of settings, and with a variety of others. But having to design their own schemas without knowing the specific supporting capabilities to be found in other schemas with which their own might be used, they are inevitably forced to create structural metadata capacities of their own within their schemas. The designers of metadata schemas (structural or otherwise) within the digital library community have sought to adhere to a particular set of design practices, seeking to create flexible, extensible, modular and generalized tools, and to promote like practice in others through inscription of their view of appropriate XML design within their technological artifacts. Unfortunately, promoting such good practices has been a death blow to one of the principle reasons for adopting XML in the first place: to ensure interoperability of digital library materials across systems. Wide-scale interoperability requires wide-scale adoption, but the design practices of schema implementers intended to promote wide-scale adoption run directly counter to wide spread interoperability.

Hence XML’s similarity to a rope. Like a rope, it is extraordinarily flexible; unfortunately, just as with rope, that flexibility makes it all too easy to hang yourself.

**Strategies for Interoperability in a World of Multilingual Markup**

The digital library community seems to face a dilemma at this point. Through its pursuit of design goals of flexibility, extensibility, modularity and abstraction, and its promulgation of those goals as common practice through their inscription in XML metadata standards, it has managed to substantially impede progress towards another commonly held goal, interoperability of digital library content across a range of systems. How then, should the community respond?
One possible response to this situation would be to say that perhaps our community cares less about interoperability than we thought. Despite projects intended to promote interoperability, such as the Digital Library Federation’s Aquifer, it may be that interoperability is actually a lower priority for the digital library community than it likes to believe, and the adoption of metadata standards that impede interoperability is merely a reflection of that underlying reality, and not a major problem to resolve. There is at least some reason to suspect this may be the case. Research libraries typically have a clearly defined local clientele, and while voices within the digital library community have been calling for some time for the liberation of content from local silos to enable their use by a larger community [Seaman 2003], libraries’ primary responsibility will always be to their local communities. The first sentence in the mission statement for the University Library at the University of Illinois at Chicago exemplifies the priorities present at most research libraries: “The University of Illinois at Chicago (UIC) Library strives to meet the information needs of UIC students, faculty, and staff.”[6] Prioritizing service to the local community is endemic to the social structure of library systems, and if systems developed to deliver digital library content to that community are successful in that context, and if the costs associated with achieving much more widespread interoperability are high, then many libraries may decide that interoperability, while desirable, is a goal which may have to wait.

If libraries do wish to make progress on the issue of interoperability of structural metadata, they will need to recognize that, as [Renear & Golovchinsky 2001] observed, “every significant information processing standardization effort must skillfully negotiate competing and apparently irreconcilable objectives, [and] serve a wide variety of stakeholders with many different interests.” In the case of structural metadata, the particular competing objectives that the digital library community does not seem to have successfully reconciled to date are what [Kendall 2007], in a discussion of blogging practices, has labeled the problem of “control vs. connection.” The structural metadata standards which have been developed to date, with their emphasis on flexibility, extensibility and modularity have sought to afford local institutions the greatest degree of control possible in their encoding practices. The standards are designed to allow any given institution to do what it wants. This has clear benefits in terms of easing adoption of the standard in any given context, and as a result insuring the standard’s widespread adoption (obviously a good thing in a standard). However, increasing the amount of local control over the ways in which a language is used and developed is fundamentally at odds with a language’s ability to serve as a means for connection with others outside the local context. It is, in essence, promoting the development of regional dialects at the expense of mutual intelligibility. The particular case of structural metadata standards reveals that sufficient local variation in syntax, the ways in which people structure their objects using a markup language, can be as fatal to communication as variation in semantics.

Given this fundamental tradeoff between internal control and external connection, libraries wishing to promote interoperability of digital library content have two possible strategies. The first, and most obvious, is to attempt to alter the balance currently struck between connection and control to more strongly favor connection. There are several mechanisms which the library community might employ in pursuit of this strategy, including the design and use of schemas which more significantly restrict both the means for recording the structure of objects and the ability to employ arbitrary additional schemas within instance documents (or developing profiles of existing schemas to achieve the same ends), establishing formal rules of structural description (equivalent to rules of description used in cataloging for creating bibliographic records) dictating aspects of object encoding not susceptible to enforcement through XML’s validation mechanisms, and mandating the use of particular controlled vocabularies and ontologies within document instances to record information such as a div element’s TYPE attribute in METS.

Decreasing the possibility for local variation in encoding of structural metadata will certainly help improve digital libraries’ capability to interoperate with each other. However, removing local capacity for variation will also tend to reduce the number of institutions who are willing to use such a markup language. If the digital library community, for instance, was to revise the METS standard to forbid any use of a relation or RelatedItem element in a descriptive metadata section to express the logical structure of a work, it would assist in insuring interoperability of digital content, but it might also very well mean losing the Library of Congress’s support for the standard. More importantly, however, such an approach overlooks one of the fundamental realities of the web environment: communities of practice no longer operate in isolation from each other (if indeed, they ever did). Even if libraries could agree on a structural metadata standard that enabled a significantly greater degree of support for interoperability than we find with today’s standards, libraries must now interact with a variety of other communities (publishers, museums, archives, educational technology companies, etc.) that are also creating their own structural metadata standards. This is not to say that pursuit of this strategy is futile or even inappropriate in many instances; libraries’ previous experience with standard efforts such as MARC
demonstrate that with sufficient time and effort a particular community of practice can achieve widespread interoperability of metadata. However, the library community's interactions with other communities clearly indicates that this strategy by itself is insufficient to resolve the interoperability problems that libraries confront today.

To deal with these wider issues of interoperability, the library community must adopt a second strategy based on accepting that the need for community control over encoding practices is a valid one, that community "dialects" of markup languages are inevitable, and that we must find ways to facilitate information exchange across the boundaries of different communities' markup vernacular. However, this will require a significant shift in the digital library community's relationship to the notion of standards. Specifically, the library community needs to shift from its current singular focus on schema development to a dual focus on both schema development and translation between schemas.

This is certainly not the case today, as can be seen if we examine the work of some of the major agencies involved in metadata standardization in libraries such as the Library of Congress. The Library of Congress currently serves as maintenance agency for a variety of XML standards developed within the library community; if you examine the list of standards that they are maintaining [Library of Congress 2008], however, you will find that while there are several metadata standards listed, standardized stylesheets to enable conversion between formats are not listed here. Such stylesheets do exist in some cases. The Library of Congress has, for example, provided stylesheets to enable conversion of MODS descriptive metadata records into MARC/XML format and back. These efforts to formalize prior work that established crosswalks between different descriptive metadata standards are not, however, seen by the community as having the status and importance of standards, as exhibited by their omission from the "Standards at the Library of Congress" web page. If the digital library community wishes to support interoperability while simultaneously affording institutions localized control over encoding practices, that situation needs to change. We can no longer view the creation of translations between standard formats as an ancillary activity; instead, we must regard it as a form of standards activity in its own right, as important, if not more important, than the creation of schemas for metadata sets.

A heightened emphasis on standardizing translation between markup languages will mean further work on formalizing translations between markup languages using XSLT, and treating those with the level of attention and care that the community has lavished on metadata schemas. However, it might also be worth considering whether the notion of formal rules of structural description mentioned earlier might be of benefit in trying to achieve greater translatability between different markup languages. As an example of what this might mean, consider the example of the 1:1 principle in Dublin Core [Hillmann 2005], that a single Dublin Core record should describe one and only one resource. The 1:1 principle provides guidance on the relationship between a metadata record and a described resource that is applicable outside the realm of Dublin Core; in fact, several other descriptive metadata standards developed since Dublin Core refer to the 1:1 principle as a guide to usage. We could easily envisage similar principles being developed for structural metadata that could guide usage of a variety of different structural metadata standards, and by working to insure similar use practices, would help insure ease of translation between different structural markup languages. We might, for instance, take as a working principle that any given structural metadata document should never contain more than two levels of structural hierarchy. Our METS example above passes muster with this rule; if, however, we modified it so that a third level of div elements was needed (of TYPE "subchapter," for example), then we would be in violation of this principle. To fix this problem, we could employ METS' mptr element to allow the div elements for each chapter to reference separate METS files containing the structural descriptions for the individual chapters. Through the establishment of common principles of structural encoding and standardized stylesheets for translation, we might be able to improve our ability to interoperate while simultaneously retaining some flexibility for local encoding practice (although obviously adoption of common principles of structural encoding may impede local control in favor of connection to some degree).

The rise of the network information society is presenting libraries with a variety of new challenges. Perhaps the most significant of these is the heightened degree of interaction with communities of practice that do not share libraries' standards, practices or values. If libraries are to survive and thrive in this new information society, they must alter their own value structure to prioritize communication with other communities to an equal, if not greater, extent than internal communication between libraries. If they pursue this course, they may find that issues of internal interoperability of library systems are more tractable than they have appeared to date.

Notes
As a reviewer for this article noted, the meaning of interoperability is vague and its interpretation highly variable depending on the community and context in which it is used. Within the discussions among the members of the digital library community upon which this article is based, interoperability is interpreted primarily as the ability to reuse data and metadata outside of the technological system in which they were originally instantiated as well as outside of the community they were originally intended to serve. Interoperability thus has both a technological and a social component.


I am including within this set the Metadata Encoding & Transmission Standard (METs, see http://www.loc.gov/standards/mets/), the MPEG-21 Digital Item Declaration Language (DIDL, see http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm), the Open Archives Initiative Object Reuse & Exchange standard (OAI-ORE, see http://www.openarchives.org/ore/) and the Fedora Object XML specification (FOXML, see http://www.fedora-commons.org/documentation/2.2.2/userdocs/digitalobjects/introFOXML.html). There are a variety of other structural metadata standards that are of particular interest to the digital library community, although not being developed within it, including standards for data archiving such as the XML Formatted Data Unit specification (XFDU, see http://eindbad.gsfc.nasa.gov/xfdu/). Being developed by the Consultative Committee on Space Data Systems and standards for structuring content for e-learning systems such as the IMS Content Packaging specification (IMS-CP, see http://www.imsglobal.org/content/packaging/) developed by the IMS Global Learning Consortium and the Shareable Content Object Reference Model (SCORM, see http://www.adlnet.gov/Technologies/scorm/default.aspx) developed by the United States Department of Defense.

See http://www.loc.gov/marc/marcsgmlarchive.html for information regarding the development of an SGML version of the MARC 21 record format.

Complete MODS and METS records for this example can be found at the Library of Congress webpage for the 1946 Library of Congress recital at http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200003790/default.html

For the complete mission state of the University of Illinois at Chicago Library, see http://www.uiuc.edu/lib/about/libmission.shtml

Works Cited


On Creating a Usable Future

JEROME MCGANN

There is a kind of anarchy that fosters growth and there is a kind of anarchy that prevents growth, because it lays too great a strain upon the individual.


I

If there is a crisis now in the humanities, as many seem to think, it is only partly about processes of tenure and promotion and ways of evaluating new forms of scholarship. The more central problem is the sustainability of born-digital resources and the work they support: the sustainability of specific undertakings (like The Grub-Street Project and The Electronic Enlightenment) but also, and perhaps even more important, the imperative to establish online scholarship, both its research and its publication, as a general institutional practice. The National Digital Public Library (NDPL) initiative provides an occasion to consider the situation of the humanities in the digital age.1

In October 2010 some forty persons from research libraries, universities, and a variety of cultural institutions and foundations met for two days at the Radcliffe Institute in Cambridge to discuss the feasibility of creating a nationwide digital library and educational archive. The meeting

The author is John Stewart Bryan University Professor of English at the University of Virginia.

PROFESSION © 2011 by The Modern Language Association of America 182
included six sessions that explored the possible content of such a library, the legal and administrative issues the library would have to address, the financial and technical demands, and the various persons and partnerships that would have to coordinate if it were to be realized. The discussions exposed the formidable obstacles to an ambitious project of this kind, but in the end the group agreed that such a library could and should be created, which it endorsed with a collective agreement to work together toward the creation of a Digital Public Library of America (DPLA)—that is, an open, distributed network of comprehensive online resources that would draw on the nation’s living heritage from libraries, universities, archives, and museums in order to educate, inform, and empower everyone in the current and future generations. By mid-November a plan was formulated to run a series of focused workshops in early 2011 that would involve a larger community of stakeholders. The workshops would be charged with developing specific institutional and technical mechanisms, funding, and work plans for the staged development of the DPLA.

An endeavor of this kind has been discussed and imagined since at least the early 1990s, and in certain global communities—for instance, China, Australia, France, the Netherlands, Japan, Norway, and Finland—national and even transnational digital archives are already under way. An interoperable world network of such libraries, an aggregated archive of archives, would be a formidable educational resource as well as an event of world-historical importance.

Of course different communities—national, international, intranational—will want to exploit and constrain the use of such resources very differently. Who would have access to what, and how could the materials be engaged and perhaps repurposed? Those overriding general questions have to be answered in practical terms that are relevant for specific persons and groups. The educational community is only one group and is itself composed of many subgroups with special interests and needs. As everyone at the Radcliffe meeting understood, the success of the NDPL venture depends on its ability to engage with those various overlapping, and sometimes competing, special populations.

Because the library, particularly the research library, is the center of all our work, we in the humanities have a particularly large stake in the success of the NDPL: success for “our cultural commonwealth” at large and not just for our special interests as postsecondary scholars and teachers. We have a crucial role to play in helping design the NDPL because of our scholarly vocation and our educational mission in society at large.

But for a scholar like me, it is sobering, sometimes dismaying, to think how much of the fundamental early work establishing the conditions for
online humanities research and publication—*OCLC, TEI, Project Muse, JSTOR*—was done by librarians, systems administrators, and various kinds of computer technologists. Regular faculty members from traditional humanities departments were not players. Indeed, to this day our scholarly community continues largely to hang back, reacting to the rapidly changing scene rather than working to shape policy and exert control over events. As a consequence, digital and Internet technology developed in their formative years at a strange diagonal to the traditional work of humanities scholarship and education.

But now many more humanities scholars are coming to participate directly, and even collaborate institutionally, in policy-shaping projects. This fact was showcased in the recent Mellon-sponsored Shape of Things to Come conference at the University of Virginia. Participants held wide-ranging discussions of a diverse set of online humanities projects, looking at the projects’ functional designs, histories, and significance for the future.

In the near future, digital technology will supplant print-based technology as the medium of scholarly publication. This is not only a fact, it is a very good thing for scholarship and education, however awkward and painful the transition we are experiencing. Print on demand itself will persist only so long as digital reading devices and, even more important, interpretive software remain as primitive as they are today. For many basic kinds of scholarship, paper, paper journals, and books are still the superior technology, though I would not hazard a guess for how long this will hold. Indeed, the launching of a venture like the NDPL by a group of scholars and their institutional fellow travelers is eloquent testimony that the ground has shifted dramatically. Until recently, such a project could live only as an idea: as the poet mused, “something longed for, never seen” (Wordsworth [*To the Cuckoo*]). The coming of Google Books was perhaps the unlooked for signal that the humanities community had a rubicon to cross.

As the crossing happens, it’s important to remember that IT has not altered the fundamental mission of the humanities: to preserve, monitor, investigate, and rethink our cultural inheritance, including the various material means by which it has been embodied and transmitted (see McGann, “Our Textual History”). The emphasis necessarily falls on the textual archive—not because libraries are more important than museums but because alphanumeric textuality has supplied most of the analytic mechanisms for organizing and studying all our material culture. That textual condition of scholarship and education will remain indispensable, but we are clearly beginning to see new interpretive possibilities emerge from
computerized information processing on one hand and graphical interface design on the other.

II

Humanities scholarship was—and still precariously is—created and sustained through the interoperation of four institutional agents. Three are structurally foundational: the scholars themselves (working in their educational and professional networks); publishing entities, especially university presses and professionally authorized journals; and libraries and depositories, where this work is collected and made accessible for reflection and repurposing. In addition, scholarship draws crucial financial and resource support from various public entities and private foundations.

Until about thirty years ago, that interlocked network functioned reasonably well. But a number of causes—the emergence of digital technology was one—began to undermine its operation. As everyone is well aware, what we now call the crisis in the humanities is a direct consequence of that systematic institutional disfunction. So in a time like ours the question surfaces: What do scholars want? Whether we work with digital or paper-based resources or both, our basic needs are the same. We all want our cultural record to be comprehensive, stable, and accessible. And we all want to be able to augment that record with our scholarly contributions.

Those desires lead many of us—indeed, most of us—to cherish the reliabilities of print-based research and traditional publication, especially monograph publication, and to resist moves toward digital venues. Alas, one might as well hope for a global economy of sailing ships or the resurrection of the Holy Roman Empire. Book culture will not go extinct: human memory is too closely bound to it. But no one any longer thinks that scholarship, our ongoing research and professional communication, can be organized and sustained through print resources.

This problem has two faces: how to pursue scholarship into a future that will be organized in a digital horizon and how to integrate our paper inheritance in that new framework. Remember, or rather (as another poet suggested) “remember and foresee”: henceforth our cultural inheritance will always be both hard-copy and digital. Even were we to digitize the whole of the world’s cultural records, even were scholars to communicate entirely in digital instead of in paper-based media, the originals would remain as important and interesting—as indispensable—as they ever were. After we digitize all the books, the books themselves will still be there.

A sharp institutional contradiction ensues, for whereas scholars want to preserve and integrate their print work for digital emergence, they also
see the need to replace print-based forms of scholarly intercourse with born-digital forms. This replacement means migrating the scholarly print archive—journals and publishers’ backlists—and also beginning to shut down the operating system of print-organized scholarly research and communication.

I say “beginning to shut down” because this system is not a machine we can easily turn off. It has such a long e-volution of its own that it is deeply integrated in every aspect of our scholarly institution. Jobs, promotion, tenure, and the entire infrastructure of the research community remain keyed to it. So we talk about prying ourselves free by shifting criteria for scholarly advancement from monograph to periodical work, or we plead that digital work—some of it, anyhow—be put on an equal footing with print work in considering scholarly merit. But as Our Lady of the Flowers said to her judge, we’re already beyond that—though not, as we all know, at the level of institutional politics.

Certainly Kathleen Fitzpatrick is already beyond that, as I think many if not most younger scholars tend to be. Her book Planned Obsolescence grounds its various proposals around a pair of premises: that “scholarship is about participating in an exchange of ideas with one’s peers” and that the traditional “system surrounding [the] production and dissemination” of this exchange “has ceased to function” in reliable ways. She is confident that we have the technical means to reconstruct this “system” in digital forms. But the charged polemic of her book reflects her worry “whether we have the institutional will to commit to the development of the [digital] systems” that will replace the “entrenched systems that no longer serve our needs.”

Fitzpatrick is an adventurous and energetic voice, and the practical cast of her mind is refreshing. But plans for institutional changes that can actually be implemented need to rest on a comprehensive view of the scholarly scene. As Fitzpatrick says, “To the degree that scholarship is about participating in an exchange of ideas with one’s peers, new networked publishing structures can facilitate that interaction.” She adds, correctly, that the interaction will work “best . . . if the discussion is ongoing, always in process.”

But implicit in those comments is a presentist view of scholarship that is alien to the humanities, where Plato and Heisenberg, Sappho and Dickinson, Scaliger and McKenzie have all been peer-reviewed. Humanities scholarship is rooted in the past, in our theaters of memory, even as it is executed in the present with a view toward creating a usable future. That is why the crisis in the humanities is only partly about tenure, promotion,
and the obstacles to a current exchange of ideas. It is about sustaining what Raymond Williams, a great scholar as well as a great critic, might have called the long e-volution if he had approached the study of culture from a scholarly rather than an interpretational vantage and if he had addressed it from our electronic perspective.

A long view is what scholars have traditionally taken and is still what scholars want or what they ought to want now. Reflect on the short view that pervades much of the thinking about (and practice with) Web 2.0, social software, and interactive online environments. Is Web 2.0 simply “a piece of jargon,” as Tim Berners-Lee has mordantly remarked? Because the roots of social networking are in online practices like Flickr and other folksonomies, the considerable scholarly potential of collaborative technology remains a pursuit.

Social software technologies have a wide-spreading but shallow root system whose most impressive result to date, Wikipedia, illustrates both its capacities and its limits. The wiki initiative delivers an encyclopedia of information that can rapidly update the range of the site’s entries and their content. How to enlist this technology for more substantial scholarship is often speculated about but not yet realized. That is to say, while we certainly have projects that implement collaborative scholarship—NINES is as good an example as any (Networked Infrastructure for Nineteenth-Century Electronic Scholarship [www.nines.org])—none of these projects is deeply integrated into the scholarly community at large. The Electronic Enlightenment, Hypercities, The Homer Multitext Project—these and initiatives like them, while open and collaborative in various ways, are still fundamentally stand-alone operations. They are crucially limited by their weak relation to scholarship’s institutional ethos—the habitus of our educational lives. Because Internet ecology at present is volatile and promiscuous, it encourages individual initiatives and just-in-time collaborations rather than programmatic strategies. This happens because Internet culture has yet to map itself to the complex social system that powers scholarship and education. So while Wikipedia, professional electronic discussion lists, and Twitter are pervaded by what Fitzpatrick calls “institutional will,” higher-level online research work—there is now a good deal of it—is not.

Fitzpatrick wants to see an “institutional will” for digital scholarship pervade her profession. This is happening slowly and tentatively, but not, perhaps, because ours is a slacker community. This slow development surely reflects the fundamental humanist commitment to the long view of culture: that a usable future is a function of a usable past. The global
character of online ecologies often casts a false appearance of pedantic narrowness on traditional scholarship. But scholarship’s investment in a long view was well established before the emergence of the Internet. The volatile character of digital resources is not just a technical matter, it is a social and ideological condition. So scholars hesitate in taking them up. Their hesitance, like Ahab’s precipitance, has its humanities.

A personal anecdote seems to me pertinent here. I spent eighteen years developing The Rossetti Archive and its content. This was a collaborative project supported by the University of Virginia’s library and the Institute for Advanced Technology in the Humanities (IATH). It involved some forty graduate students plus a dozen or more skilled technical experts, not to speak of the cooperation of funding agencies and scores of persons around the world in many libraries, museums, and other depositories. It now comprises some seventy thousand digital files and forty-two thousand hyperlinks organizing a critical space for the study of Rossetti’s complete poetry, prose, pictures, and designs in their immediate historical context. The archive has high-resolution digital images of every known manuscript, proof, and print publication and of every known or accessible painting, drawing, or art object Rossetti designed. It also has a substantial body of important contextual materials. All this is embedded in a robust environment of editorial and critical commentary.\(^1\)

I undertook the project partly as a laboratory experiment to explore the capabilities of digital technology and partly to create a scholarly edition of Rossetti’s work. As a laboratory experiment the project was a remarkable experience—a clear educational success, I should say. I used to measure that success in theoretical and intellectual terms—as indexed in the series of books, lectures, and essays that spun off those eighteen years of the archive’s development. I now measure the success by the project’s institutional position and relations: where it came from (the various digital initiatives at the University of Virginia) and what it led to (Speculative Computing Laboratory, the software development group Applied Research in Patacriticism, and NINES).\(^2\) Most important are those young men and women, then graduate students, who are now among the generation of scholars shaping the future of humanities research and education.

On the other hand, if the archive is judged strictly as a scholarly edition, the jury is still out, for one simple and deplorable reason: no one knows how it will be or could be or even if it should be sustained. The supreme irony of this adventure: I am now thinking that, to preserve at least the core of its scholarly materials, I shall have to print it out. (It would extend to many volumes.) I have also come to think that the archive’s most important content is nothing digital at all.
III

_The Rossetti Archive_ and projects like it are most important because their historical backwardness is beginning to show. Their own process of development has exposed the social and conceptual limits of the digital ecology that spawned them. These limits, which lie concealed by the (often) impressive appearance of such works, were shaped by historical and institutional circumstances. As we try to map our way to the usable future promised by the NDPL and by an even more comprehensive world library, we need to understand those conditions.

The early history of IATH, where _The Rossetti Archive_ and similar projects were developed, is exemplary for the issues we now face. For a dozen or so years beginning in 1992, when IATH was founded, the institute was a focus of keen attention for humanists interested in digital technology. Part of the interest was in the institute’s groundbreaking projects—_The Valley of the Shadow, The Rossetti Archive_—and part was in its strategic commitment to high-end humanities research work.

IATH’s founding modus operandi was based on a set of four explicit working premises:

To organize its institutional operation as a freestanding unit accessible to all UVA faculty members but not answerable to departments or deans
To seek out and exploit the interests and strengths of particular scholars within the university
To promote intra- and extramural research projects (rather than classroom or pedagogical projects), on the assumption that in postsecondary education important pedagogical work is a function of important research work
To design all projects for a global Internet environment

IATH’s successes were shaped by those directives. With research agendas at the center of the institute’s work, IATH began to explore the digital humanities as an autonomous disciplinary practice. Establishing the institute as an independent unit set it outside the direct control of the university’s divisional and departmental authorities. It was judged—correctly—that the faculties at large were not ready to promote the institute’s work in programmatic ways. At the same time, the institute actively sought out and supported individuals in the faculty who were committed to pursuing digitally based high-end research work. Allegiance was pledged to the belief, long held by the university community, that innovative research would drive effective and innovative pedagogy.

Finally, designing the institute’s projects for the emerging Internet set all the work in a global context. That orientation had profound consequences.
We were strongly biased toward free culture and open access, toward non-proprietary software and open-source development, and toward a commitment to distributed networks and generalized standards for metadata and text encoding. So the initial premise for designing the logical structure of *The Rossetti Archive* was that a distributed Internet archive of archives already existed. Given the emergence of such a network in the future, and understanding the uncertain directions that hardware and software developments would be taking, what kind of design would most ensure that *The Rossetti Archive* would get integrated into that foreseen but unrealized situation?

When I posed that question in 1992—even in 2002—I conceived it as a question about the formal design of an information system for scholars. On one hand, we sought to construct a logical ground for the archive’s content that would not be compromised by new software and hardware developments that were certain to come. On the other, we left the sociological face of the question to the global reach of the Internet, which (it seemed) scholars could exploit without having to get involved in its infrastructural design and development. In any case, the latter (it also seemed) was being driven by forces beyond the reach or the resources of the humanities research community.

IATH flourished in the 1990s and early 2000s because its orientation mirrored the general state of play in online humanities scholarship at large. But with the turn into the twenty-first century the limitations of the IATH model began to become clear.

Of first importance here was the burgeoning of various types of interactive social software in the early 2000s. Mapping digital scholarship to the ontologies, but not the sociologies, of the Internet not only constricts its institutional presence but also obscures the sociohistorical character of traditional philology itself. For all their “material” conditions, their hypertextualities, their Internet connections, and their collaborative features, the IATH projects had a design that was predominantly formal and abstract. In this respect they have been a mirror and model of nearly all content-focused online scholarly projects. Like *The Rossetti Archive*, *The Blake Archive* is without question the most comprehensive and authoritative edition of its subject ever created, as well as the most globally accessible. It is not, however, the edition that scholars work from or cite. Paradoxically, ironically, innovative online scholarship emerged, and for the most part still remains, peripheral to the regular research and publication of the vast majority of working scholars.

NINES was born (in 2004) as a response to that scholarly condition. Unless they are integrated into the sociologies of an online world library,
projects like *The Rossetti Archive* are only minimally useful to scholarship. So NINES was conceived as a small model for exploring the problem of informational design for scholarly work at a global scale. It developed an operational (institutional) response to the following question: assuming a distributed world network of objects like *The Rossetti Archive*, how should the network be organized and its materials integrated? This is very like the question that shaped the initial development of *The Rossetti Archive*. It differs in one crucial respect: it has a social and institutional horizon. The inquiry addresses not the formal design of a complex autopoetic system but the structures needed to promote access and repurposing by a distributed population of research scholars and educators. Most immediately, this meant that working scholars would authorize and develop the NINES content so that the traditional work of scholarship—research and publication—could be pursued in the kind of integrated environment characteristic of our inherited paper-based system. Like the latter, NINES is designed to grow and develop through the use and input of scholars who want reliable sources and trusted materials and who expect their own work to be peer reviewed.

A key initial decision, therefore, was to move against the promiscuous state of information available on the Web. The move operated on two fronts simultaneously. First, NINES established itself as a peer-reviewing agent that would identify and assemble a corpus of authoritative online materials. These would have to include every kind of online resource that a working scholar uses: stand-alone online projects like *The Walt Whitman Archive*; library and museum catalogs and accessible content; proprietary materials (like those developed by Intelex, ProQuest, and the Alexander Street Press); online scholarly journals archives, both free culture and subscribed (for example, *JSTOR, RaVon, Project Muse, 19*); and university press publications.

Second, NINES sought a technical and disciplinary structure that would permit the range of authoritative resources to be indefinitely expandable. The particular goal here was to develop a comprehensive scholarly environment—an online corpus with, for example, the MLA’s disciplinary range. So NINES from the outset was working to promote similar entities for medieval, Renaissance, eighteenth-century, and modernist scholars. The first of these, *18thConnect*, was formally launched in 2010, and the others are now in active development.

IV

Throughout the 1990s, humanists working with IT were tempted to believe that they were pioneering a new disciplinary field called the digital
humanities. The last ten years have largely changed our minds and sent us searching back to our future. Online ecologies are leading us to imagine a new philology, a critical and disciplined study of history and culture whose center could be anywhere and whose circumference nowhere. Indeed, as the global capacities of online networks have grown, so has the need to control and organize them for particular social and institutional needs. So while digital technology is introducing new critical methods and procedures, it does not fundamentally alter the sociologics of scholarship and education or their institutional mechanisms. Studying the history of philology itself is especially pertinent now, as is a broad critical reflection on the current institutional state of humanities education.

In that context, we now have a pressing need to integrate online humanities scholarship into the programmatic heart of the university. Twenty years ago, for various institutional reasons, university degree programs could not support advanced work in the digital humanities. So online scholarship flourished in extraprogrammatic localities: typically, in the orbit of the university library or through special agencies like IATH. The NDPL initiative is the unmistakable sign that we can no longer safely proceed in that way.

NDPL is a project to install a comprehensive online archive of authorized and reliable materials for general public education. Philology is the discipline for realizing—in the words of William Wordsworth (“Lines Composed a Few Miles above Tintern Abbey”)—“what [we] half create, and what perceive” as the order of history and culture. The critical study and augmentation of such an archive is the foundational mission of philology, and university degree programs are a sine qua non for executing that mission. To date, however, this cultural archive has been growing with little direct input or oversight by the community—humanities scholars and educators—who have a fundamental interest in its successful realization.

Consider the Google Books settlement. Its most disturbing aspect has been that the higher education community was not represented in the negotiations.14 But the dismal truth is that humanities scholars have been absent for years from many decisive, if less dramatic, events. We have been like marginal, Third World agents who have actually chosen an adjunct and subaltern position. We have been invisible. Because only a small minority of scholars has been active with digital work and the institutional changes it is bringing, they have been functioning on their own or in insulated venues like IATH, removed from the university’s programmatic community. This is a serious institutional fracture in the world of humanities scholarship and a major source of what has come to be called the crisis in the humanities.
The *Google Books* settlement was the historical event that triggered the NDPL initiative. “In what we now call the information society . . . we need a new ecology, one based on the public good instead of private gain” (Darnton). These words show us how the NDPL ought to be an invitation and an opportunity for humanities scholars. Realizing the idea of the NDPL means developing the institutional means for reorganizing and reediting our entire cultural inheritance, traditional as well as digital, for online access, reflection, and repurposing.

A library of this kind is so central to the educational mission of the humanities that we must insist on helping shape its future. As it emerges, humanists will also have to begin promoting the digital transformation of postsecondary humanities programs and curricula. The institutional inertia that has been resisting these changes, while often deplored by digital enthusiasts, can and should be rethought and redirected. Traditional scholarship operates through a complex machinery of paper-based social software about which digital technicians are often deeply ignorant. The ontologies needed to organize an effective online educational system are already operating, largely transparently, within the social network of traditional scholarship and education. The machine of the book, perhaps the greatest social technology ever invented, should be the object of deep study by human-interface designers. But of course that kind of study, the repurposing of the work of the past, is exactly what we mean by scholarship. *The Blake Archive* is a repurposed migration not only of Blake’s original works but of all their subsequent scholarly migrations and transcriptions, and most especially the foundational editions from Ellis and Yeats to Keynes, Erdman, and Bentley.

“We will advance funeral by funeral,” a learned digital scholar once mordantly remarked when I was kvetching with him on these subjects. And while I’m sure he touched an important truth, it isn’t a truth to help us shape reliable policy, which is what we need. Sustaining digital scholarship means sustaining our cultural resources *tout court*, digital and nondigital, and it also means taking a long view. It is a social problem pressing on the entire community entrusted with the care of public education. Advertising, ideology, propaganda, and entertainment are part of our public education, but scholarship is its source and end and test.

NOTES

1. At this writing, the name for this initiative is still being discussed. It began as NDPL but then became DPLA (Digital Public Library of America). It may change once more.

2. From the steering committee’s summary of the sense of the initial meeting, sent as an e-mail message to the participants.

4. A larger community of stakeholders was brought into the organizational meetings, which began in late February 2010.

5. The commission that issued the report Our Cultural Commonwealth was chaired by John Unsworth.

6. Individual scholars and small groups have initiated special, self-contained entities (like Stephen Railton's Uncle Tom's Cabin and American Culture or Jeffrey Savoye's The Edgar Allan Poe Society of Baltimore). The TEI is once again exceptional, since its mission is precisely to establish general text-encoding standards to foster project interoperability in a dispersed network.

7. A print-on-demand collection of the conference papers and responses is available at http://cnx.org/content/col11199/1.1/collection_print_confirmation.

8. When this movement is referred to as the digital humanities, the implication is that it constitutes a disciplinary subfield in the humanities. My own view is that it is not a field like philosophy or English or global studies or even book history or media studies. It is primarily critical-methodological rather than substantive, a set of investigative procedures along with the tools for executing them. In terms of literary and cultural studies, it perhaps most resembles a fundamental subdiscipline like bibliography.

9. In relation to those original objects, digital simulations are epistemological, not ontological, objects. They are acts of interpretation.

10. I quote from the free online (prepublication) text of her book. This paragraph and the next three originally appeared in my essay “Sustainability: The Elephant in the Room.” The salience of the issues raised here is slightly but significantly different from the appearance of the issues in “Sustainability” and in the volume of essays that it introduces.

11. The archive is a complete collection of all Rossetti's textual, pictorial, and design works in all his textual works, in all their known material forms and states. There are 845 textual works that exist in some 14,000 distinct documentary states and more than 2,000 pictorial and design works. Each document has an xml transcription as well as a high resolution image, and with a few exceptions each artistic work is represented by a high resolution image of both the original work and, in many cases, various later important reproductions of the original. In addition, the archive has some 5,000 files of extensive scholarly commentaries and notes on its materials.

12. For IATH, see McGann, Radiant Textuality; for Speclab, see Drucker; and for ARP, see McGann, Applied Research.

13. Although the history of social software can be tracked to the earliest days of IT, the period 2002–04 is widely recognized as especially significant, because in 2002 Clay Shirkey held his invitation-only “Social Software Summit,” and in 2004 Tim O’Reilly convened the first of his annual Web 2.0 conferences.

14. At least not until Pamela Samuelson’s critical interventions that culminated in her letter to the court of February 2010. See Samuelson.

WORKS CITED

SCHOLARLY COMMUNICATION INSTITUTE 9

New-Model Scholarly Communication:
Road Map for Change

University of Virginia Library

July 13-15, 2011

Abby Smith Rumsey, Director
Scholarly Communication Institute

INTRODUCTION AND MEETING SUMMARY

The Scholarly Communication Institute 9 (SCI 9) convened scholars, librarians, publishers, higher education administrators, and funders to develop collaborative strategies that will advance humanities scholarship in and for the digital age. SCI participants explored the impact of trends within and beyond the rapidly evolving landscape of higher education on scholarly production and communication, from producing and publishing to assessing, curating, and using. They examined the new roles and responsibilities assumed by the major actors in scholarly communication. They proposed actions to create sustainable infrastructure built on shared capacities and addressing shared needs. And they identified the people and organizations that are best positioned to play leadership roles in modeling, validating, and sustaining new-model scholarly communication.

This report synthesizes discussions among participants at the ninth and last session of the Scholarly Communication Institute at the University of Virginia. This culminating institute built upon insights gleaned over nearly a decade of meetings and programs sponsored by SCI that were designed to model and test new approaches to humanities scholarship. Each institute focused on distinctive aspects of new communication strategies and modes of working as they refashion humanities inquiry within disciplines (practical ethics, architectural history); through the use of new technologies (visual and geospatial technologies); in sites of innovation and experimentation (humanities centers); and, at SCI 8, in emerging genres of discourse that take advantage of digital affordances. This final session took an integrated look at how multiple
actors in the landscape can interact and collaborate to accelerate desirable change across the scholarly communication system as a whole.

What is new-model scholarly communication? By scholarly communication we mean the authoring, publishing, stewardship, and use of scholarship. Digital scholarship is the use of digital evidence and method, digital authoring, digital publishing, digital curation and preservation, and digital use and reuse of scholarship. And new-model scholarly communication is what results when we put those digital practices into the processes of production, publishing, curation, and use of scholarship. The goals of scholarly production remain intact, but fundamental operational changes and epistemological challenges generate new possibilities for analysis, presentation, and reach into new audiences. The changes also pose serious challenges to existing organizations, professions, and business models. These developments are unfolding in an era of radical public-sector defunding across higher education, the erosion of tenure-track career paths, profound organizational disruptions in scholarly associations, libraries, archives, and museums, and the break-down of print-based publishing business models.

Advancing the humanities in and for the digital age demands the active engagement of many sectors of the scholarly community working towards a shared vision. The key actors in the successful transition of humanities to a digital environment are:

- **Peer communities** of scholars able to assess and validate new forms of scholarship, including genres that cross disciplinary boundaries, reach new audiences, and use technology in innovative ways
- **Publishers** able to support new communities of discourse producing scholarship in multiple media and genres, and engaging the attention of diverse audiences
- **Libraries** that can support use of sources in all formats and ensure the integrity and long-term accessibility of the scholarly record
- **Administrators** who enable realignment of resources to where they are needed and support changes in the reward system of scholarship
- **Funders** who seed innovation through projects designed to model and test new modes of working, new methodologies, and new career paths for the many professionals involved in scholarship.

Participants pointed to evolving practices and mindsets that take advantage of digital affordances to push humanistic inquiry into new territories. They homed in on a set of collective priorities and specific venues to serve as common ground for aligning interests and resources. They proposed a series of actionable ideas, detailed in the following report, in five areas:

**Develop a shared mental map for scholarly production and communication in the digital age.** Improve and normalize the processes and workflows of digital scholarly production and authoring; and connect with and sustain new (and larger) audiences in addition to peer disciplinary audiences
**Reengineer the system of credit.** Explore and articulate criteria for assessing scholarly merit in the online environment; experiment with venues for peer review to increase transparency, reliability, and participation; devise methods to sift through the surfeit of available information and direct scholarly attention to meritorious work; and realign reward and recognition systems to apportion credit where credit is due.

**Nurture new career tracks and provide better professional development.** Define new career tracks; provide education and training for new skills and literacies as part of graduate education and throughout the working life of humanities professionals.

**Better align shared capacities to attain sustainability throughout the lifecycle of digital scholarship.** Scholarly societies should reconceive the core services they offer to their disciplines and members; publishers, libraries, and museums need to develop new business models that meet the aspirations of open access to content; and libraries and museums should retool their investment and policies to ensure the continuity of their stewardship responsibilities while scaling up for significantly greater curation and preservation roles in the digital age.

**Engage new sources of support and funding for the humanities.** Target strategic areas for funding, such as building consortial models, stimulating experimentation, and supporting professionals at critical transitional phases in their development.

**NEXT STEPS**

Digital practices are remaking and redefining humanities in and for the digital age. As the humanities migrates with the society it studies and serves into online communities, scholarly communication professionals will be far better positioned to pioneer digital practices that simultaneously create and demonstrate value to society.

The following is a summary list of recommended near-term actions and the actors taking leadership roles in implementing them. The full report that follows provides rationale, context, and details. The professions that are responsible are listed in brief, and should be understood to include both individual organizations such as scholarly societies, and umbrella organizations such as the American Council of Learned Societies (ACLS), the Association of Research Libraries (ARL), and the American Association of University Presses (AAUP).

**Scholarly Production & Authoring:** continue numerous ongoing experiments in new platforms for publishing and new genres (Scalar, Press Forward, MediaCommons and MLA’s Digital First); develop metrics for use, influence and impact, and lasting value to scholarly discourses (scholarly societies, humanities centers, publishers, libraries, funders)
Assessment & Credit: articulate benchmarks of scholarly merit in digital scholarship; undertake scholarly network analysis; develop readily adoptable peer review and commenting systems for post-publication assessment; develop assessment and “microcrediting” systems for collaborative work involving professionals in scholarly communication (scholarly societies, humanities centers, libraries, publishers, administrators)

Shared Infrastructure: develop partnerships among scholars, libraries, and publishers to support new, streamlined production-and-use workflows that operate throughout the lifecycle of digital creation (scholarly societies, libraries, publishers)

Education & Professionalization: revise curricula to include skills currently required in scholarly communication professions, such as digital research and development methods, project management, design and editing skills, public writing and speaking; increase numbers of and access to venues that provide continuing education in new-model scholarly communication, such as regional institutes, and distance learning courses; investigate the appropriateness of the dissertation as presently practiced for preparing graduate students either for a lifetime of sustained scholarly productivity or for other intellectual but non-professorial career paths (scholarly societies, humanities centers, publishers, funders)

Funding & Support: develop a compelling articulation of how and why humanities is crucial for the digital age; and engage new sources of support for digital collection building, professional development, library and curatorial skills (SCI, ACLS, Council on Library and Information Resources (CLIR), funders)

The list of participants, agenda, background materials, summary of pre-session questionnaires, and additional materials generated during SCI 9 can be found at: http://www.uvasci.org/. The following essay reports on and integrates SCI debates on the near-term opportunities for moving the full cultural record of creation, reflection, and meaning-making online to shape a new information environment that embeds the histories and values of the humanities.

*****

THE CHANGING LANDSCAPE OF HUMANITIES AND SCHOLARLY COMMUNICATION IN THE DIGITAL AGE

This capstone session called upon participants from all sectors of scholarly communication—scholars and scholarly associations; organizations such as libraries, museums, archives, publishing houses, humanities centers, and professional associations; and higher education administration and funders—to
articulate a common agenda of actions necessary to strengthen “humanities in and for the digital age,” as historian William Cronon phrased it. Although we invoked the metaphor of a roadmap for change, the image may mislead by creating a reassuring sense that we know where we want to go, that we will know when we get there, and that there is, indeed, a road to follow. Given the pioneering spirit exemplified by the participants, it is more apt to think of moving through uncharted territory and clearing a path for the humanities in a changing and unpredictable environment. The working mode of SCI has been to engage the imagination—we ask what an ideal system would look like if designed from the ground up—and realize that vision by modeling, testing, and then normalizing new-model scholarly communication strategies in everyday practice.

TRENDS IN HUMANITIES SCHOLARSHIP

The first task of realizing this vision is to inventory existing ambitions, capacities, divisions of labor, and gaps within institutions and peer communities. Participants examined inherited assumptions about the system of scholarly communication and identified trends now shaping knowledge creation and sharing both within higher education and beyond. Digital technologies are everywhere pushing against traditional practices from searching to citing, and from writing to reading. Significant trends within higher education to be reckoned with include:

• disruption of the academic job market: “adjunctification” of research and teaching faculty, loss of tenure-track positions, recombining of roles in teaching and research, post-doctoral positions with no clear career to move into, and the rise of self-styled “alt-ac” or alternative academic careers by highly skilled humanities professionals opting out of proto-tenure careers and into new-model scholarly professions

• disruption of supply of and demand for expertise: shaking up received notions of professional profiles and expectations in research, teaching, publishing, librarianship, and, consequently, preparation for these careers

• disruption in the roles and responsibilities played by institutions: libraries, archives, and museums are taking on additional digital stewardship responsibilities and publishing content to the Web; digital humanities centers often exist alongside traditional humanities centers rather than within them; academic publishing houses are being challenged by new publishing consortia and enterprises; scholarly associations’ traditional services of hard-copy journal publishing and annual meetings are eroding in value

• breakdown of business models: publishers, libraries, and scholarly societies are struggling to provide services that add value and are
financially sustainable; chronic lack of access to even modest capital hampers their ability to experiment and innovate

• *breakdown between theory and practice:* forced by new communication technologies that strongly link making to theorizing, intellectual agendas are exploring new relations between media and knowledge creation that upend print-based practices of assessment and recognition

These changes have particular meaning in the context of larger social, political, and cultural upheavals in the very world that higher education is designed to prepare students for. Most relevant for the humanities are:

• *changes in the nature and constitution of the audience* (for humanities and all online information): readers now expect to be active users and producers of content, not passive receivers of information; the time span between creating and posting content is short, and reception and reaction equally short

• *radical public-sector defunding of infrastructure:* this occurs in conjunction with stresses and break-downs in analog business models and raises more fundamental questions about the value of humanities to society. The private-sector funder *Don Waters* noted massive defunding makes questions about whether the digital humanities is real scholarship seem trivial. We must “tie our work to the larger tasks of articulating, defending, and advocating the role of humanities in post-industrial, globalizing society.”

• *rise of informal peer-to-peer networks of knowledge:* the blurring of distinctions between expert and lay, academic and public scholars, and scholars and the public is potentially a sanguine development in a democracy that assumes a well-informed citizenry; but it poses challenges to professionals and the processes of professionalization

• *IP regimes:* across all media they are out of sync with the impetus to share, use, and reuse knowledge; and they fail to provide the necessary means for cultural heritage institutions to curate, preserve, and ensure long-term access to digital content

**HUMANISTIC RESPONSES TO THE DIGITAL AGE**

In the scant decade or two since the beginning of the migration of knowledge creation and sharing to the Internet, humanists have faced these trends squarely—though far from uniformly—both as a series of unfolding challenges to address (especially in areas such as IP and publishing), and as extraordinary opportunities for humanists and their disciplines to engage new questions with new generations of students and users. The growing number of humanists who are embracing multimedia methodologies and more collaborative modes of
working, such as the SCI participants, are reporting very promising developments in research, teaching, and public engagement, including:

- new links between evidence and argument in publication, deeply inflected by the availability of abundant primary sources online
- new attention to curation and aggregation of sources and interpretations
- new communities of discourse arising rapidly to share, validate, comment on, and point to new scholarship
- changing vectors of academic communication from vertical to horizontal, engendering new rhetorical forms and engaging new audiences
- new professional tracks, such as alt-ac (alternative academic), to assume new positions in libraries, humanities centers, research labs, and presses
- new service models in libraries that support collaboration, curation, and dissemination
- new sites for scholarly production, assessment, and dissemination in learned societies and humanities centers
- new alliances between traditional and digital humanities centers, based on common research agendas around disciplinary change and emerging digital publics

**GAPS & NEEDS**

What will it take to accelerate these developments? Steven Wheatley of the ACLS provocatively compared the challenge of new scholars and new disciplines to that of entrepreneurs.

Every scholarly career is something of a start-up enterprise. The academic develops a particular product—a book, an article, a topic, but really a specialization—and then brings it to market in a field or discipline. If the product finds acceptance (finds customers, that is), then the start-up is started up and the ambitious, active scholar will be able to follow newer and more developed version[s] of her product.

This trajectory holds for new scholars and new disciplines alike. What is important is bringing an idea and a product to an audience; and it takes a deep and extended educational system to create good scholars, to support their work, and to support the conversation between scholars and audiences. In the digital age, the start-up demands bringing together many dispersed and siloed strands to support scholars and audiences, to align strengths and interests, match resources to needs, and create new partnerships among actors who may be unknown to each other but share common goals.

Concerted actions in five overlapping areas—scholarly production and authoring, assessment and credit, infrastructure, education and professionalization, and funding and support—were defined and charted for our road map, based on cross-sectoral debates that ranged over theory and practice, people and places, processes and products, and action and reflection.
SCHOLARLY PRODUCTION & AUTHORING PROCESSES

Participants emphasized that scholarly production and communication are being remade not only through technologies per se, but even more so by vast changes in audience, attention, and the construction of authority. The template for print-based scholarship operative in the last century is losing value more quickly than a new template is taking its place, though, and this is creating a sense both of insistent pessimism about disruption in some quarters (though not among those at SCI) and exhilarating organizational and epistemological ferment on the other. What has emerged is still very fluid and literally under construction. We will focus here on three tested models addressing audience, attention, and authority that will serve as a lens through which to view the larger landscape of telling changes and of the shape of things to come. The models examined at SCI also shed light on the fundamental issues under debate happening on campus, online, and in face-to-face meetings: what are and should be the primary genres of humanities scholarship in the digital age, and how their scholarly merit is recognized and rewarded.

1. Multimedia Publishing

How do we fully integrate and take advantage of multimedia sources (moving and still image, audio, cartographic, textual, and manuscript data), analytical tools (geographical information systems, data mining, visualizations, sampling) and presentation strategies?

Disciplines such as visual culture, media studies, sound studies, and architectural history, among others, are embracing the potential of multimedia to represent primary sources in rich media and to present interpretations. The Alliance for Networking Visual Culture (ANVC), a coalition first conceived at SCI in 2007, has worked with university presses and media archives to pioneer a new platform for research and publishing, Scalar. (Vectors is an earlier example of such an inventive platform.) ANVC scholars are propelled by new ways of collecting, curating, and analyzing data to create publications in which image, sound, text, geospatial visualization, and so forth, are not only routinely used, but are fully integrated, one not subservient to the other. Tara McPherson, a media scholar from USC, and Ellen Faran, director of MIT Press, reported on their collaboration around Scalar and directed our attention to the recent publication of a Scalar-based work, Learning From YouTube by Alexandra Juhasz. In addition to the positive response that this new-model publication has inspired among many scholars, it brings forward a host of matters that need to be rethought in the digital environment. Multimedia publications should provide:

- Integration of primary resources into a short- or long-form argument
- Normalized citation practice. This implies a fixed version, which may run counter to the impulse of time-based and interactive media, and thus raises significant epistemological issues
- Protocols for quoting. These should enhance scholarship and teaching and help to rationalize fair use in the online world
• Accommodation of varying granularities. McPherson suggested rethinking or inventing hybrid forms, in which the short and/or interactive is tied (somehow) to a longer form

• Collaborative space within the project. This should host the entire production team as it begins to work on the product and extend to allow creators access to archives and applications such as the ANVC partners Internet Archive, HyperCities, the Hemispheric Institute, and other projects and publishing environments, scholarly societies, and humanities centers

• Copyediting. Editorial staff need new procedures and workflow for multimedia copyediting because it is “seriously different from copyediting in traditional books”

• Editorial skills that embrace design. Editors now need to have what Faran call “spatial orientation,” understanding how the multiple elements of the presentation—text, media, workspace, annotation and commentary—fit together

• Need for new user metrics. Publishers do not yet know how to gauge if they are reaching the desired audience

These needs all provoke deeper intellectual questions and spark the “exhilarating ferment” that is an abiding affect of this work. How can we use archives to make novel arguments or, as McPherson put it, “How might scholars inhabit the archive anew?” And in the case of Shoah Foundation’s historical materials about the European Holocaust, how do we pay attention not just to the audience’s needs or the creators’ desires, but to the demands, ethical and otherwise, of the sources themselves and the organizations that assume responsibility for securing them over time?

Evaluation of scholarly merit becomes problematic. People expert in scholarship and technology are needed, because the benchmarks of excellence in argumentation include the design and programming. Further, we have only a provisional sense of reception, both within the academy and beyond. Are there analytics that will help? Trying to adapt traditional book analytics here is tricky, not least, one press director said, because such metrics turn out to be more notional than real in the end, having the power of tradition behind them but surprisingly little else.

Finally, Faran asserted that a project’s team of authors, editors, technologists, designers, and marketers need to come together around these questions well before the product is written, assembled, coded, and programmed. It is extremely important that multimedia publishing is done as a collaboration between scholars and publishers, not as arms-length hand-offs from scholar to editor, editor to designer, designer to programmer, programmer to marketer. The conversation must begin at the beginning and be sustained through the iterative process of scholarly production.

2. Direct-to-Web Scholarship
How do we create and share knowledge on the dominant and most democratic publishing platform of the digital age—the World Wide Web?

The widespread phenomenon of publishing directly to the Web challenges our assumptions about writing, reading, and attending to scholarship. What Dan Cohen and Tom Scheinfeldt, historians and director and managing director respectively of the Roy Rosenzweig Center for History and New Media characterize as the Web way equates to disseminating scholarship online in multiple stages of development, inviting review and building audience simultaneously. Recognizing that the Web places novel demands on our attention, our notions of authority, and especially on building and reaching audiences, Cohen and Scheinfeldt are creating a publishing platform, Press Forward, and suitable protocols that will prototype Web-based, peer-reviewed scholarly discourse for the humanities and beyond. They aim to create new audiences for scholarship at the same time as they expand new modes and tempos of discourse for scholars and creators. Their goal is to "develop effective methods for collecting, screening, and drawing attention to the best online scholarship." In the process the builders expect to generate as many questions as answers for some time to come, thus becoming one of “the problem factories” that Waters said we need many more of at this juncture.

This model does not envision the scholar authoring in isolation, submitting the product to a group of selectors/editors, with review and changes made prior to publication. It starts where people and content are—on the Web—and garners attention, attracts an audience, and creates value as it engenders a community looking at and reacting to common things. Press Forward takes scholarship published directly online as the starting point and focuses on aggregation and curation. In this model, evaluation, critical response, and validation happen in a more open fashion that involves multiple respondents, distinct from traditional closed review by an anonymous handful of peers. Identifying value therefore means sifting through a superfluity of scholarly and scholarly information to find items that merit attention, based on peer assessment, citing, and use. The community creating and using the discourse must figure out how an evaluative body can point to it. The builders of Press Forward anticipate that this will develop into a "plug-and-play" platform with relatively low entry costs that scholarly societies can use. By doing so, the value that societies would add to content is the authority and reach of the scholarly association.

This experiment in Web-based discourse entails an extended view of the scholar as author. Scheinfeldt refers to the scholar crafting an online persona as the Web way of creating a profile and portfolio. This demands active curation and stewardship of one’s own scholarship over time. The good work habits that shape a portfolio over a career must begin in graduate school. This Web-based approach also has implications for the libraries and other stewardship organizations necessary for the long-term integrity of and access to this material. It implies the development of business models that can support production, distribution, use, and curation in environments that are essentially open and evolve over time.
The concept of demand proves to be a useful frame for fresh thinking about audience. Until recently, arbiters of demand for humanities scholarship have been university presses and curriculum committees. While still quite powerful, they may be running against a strong current and risk being caught in the drift. As the historian Jim Grossman noted, we are still trained to do long-form research and writing, but how many of us do long-form reading now? We need to develop more granularity in our genres, such as the kind of hybrid forms that McPherson mentioned. The modeling and testing of such approaches as Scalar and Press Forward are the very best way for scholars and audience to connect and discover what the appropriate modules of granularity are. But there need to be many more such robust models-as-problem-factories—sites of invention, reflection, participation, and learning. The bodies of prestige, validation, and authority such as scholarly societies, university presses, humanities centers, and individual scholarly leaders need to get involved and become responsible partners.

3. Dissertation and Monograph Genres

What are the scholarly digital genres that accommodate short- and long-form arguments, and what do those digital genres tell us about the “dissertation-as-proto-book” as the most appropriate preparation for a career of productive scholarship? Indeed, what are the intrinsic values of short and long forms in the digital era?

Concomitant with the appearance of multimedia and Web-based genres, the fate of the monograph and its training form, the doctoral dissertation, is very much under scrutiny. The first-order question here is about the relationship between disciplinary epistemologies and communication technologies. Which fields require new information technologies? Are there fields yet to emerge that we can uncover only with these new technologies, as the history of the book suggests? Something also to consider is to what extent the monograph is at risk intrinsically, or under economic pressures, or both. Finally, if we were to invent a long form for the Web, what would it look like, who would be its audience, and what preparation would we require of graduate students aspiring to create it? Chances are, people concluded, the invention would not bear much resemblance to today’s monographs. Nor would its training form be today’s “proto-book dissertation” form.

The Modern Language Association (MLA) has looked into these questions and the dissertation in particular with their New Dissertation program to explore the forms necessary for the dissertation. Do we really need to keep using the codex as a model? MLA Council member Kathleen Woodward noted the dissertation is a hinge on which so many things depend and take shape. The dissertation cannot take a form that is not economically viable (ie, the monograph); and economics here means much more than business and cost models. It means how the academy decides where the human capital of its graduate students should be directed as they prepare for a life of productive scholarship. Kathleen Fitzpatrick challenged us to think about what a life of productive scholarship
will look like over the next thirty years. The answer is “We don’t know.” But no one would bet on it revolving around putting out a succession of hard-copy monographs on increasingly specialized topics. Hence the legitimate confusion about what the dissertation should be now.

McPherson argued that what the dissertation should be is a project that demonstrates intellectual maturity, is not beholden to advisors, and works from essentially original research. Wheatley remarked that in terms of assessing a junior scholar’s output, the book and article have no inherent value; their value is an assumed future value. The dissertation is meant to demonstrate capacity in relation to some body of knowledge, so among the things the dissertation should do is demonstrate knowledge of an existing corpus and demonstrate capacity as well. Capacity for what is the question now. Woodward challenged the notion that capacity for argument should be the sole essential criterion for assessing scholarship. Is the dissertation something whereby we create and persuasively present new knowledge? Is it about the curation and preservation of knowledge? Can we imagine that a new-model dissertation would be a translation, a collection of essays, original digital objects, or curatorial projects?

Another question is how the attitudes around the dissertation change the nature of the late-stage graduate career. Wheatley reminded us that there is no longer a rush to get people into the academic job market, given the paucity of demand. As for the post-doctoral academic positions, which are increasingly seen as a purgatory that does not have a clearly marked portal to the heaven of tenure-track, there was a consensus that these positions need to be radically rethought. One thing we’ve ascertained about 21st century careers: the ability to navigate the online environment and to disseminate knowledge to an audience is critical for both academic and nonacademic jobs. Digital literacy is crucial to impart during graduate training. Senior scholars reported seeing too many cases of what Woodward called “anticipatory remorse,” when advisors will not support a certain kind of dissertation (ie, digital) because they think it will ruin their students’ career chances. How can the expectations for humanities scholars be so at odds with trends in society? (The question of what education and professionalization means is taken up in greater detail below.)

While participants define the specific end goal of the dissertation as preparation (even though that left huge questions about preparation for what), the monograph has a harder time finding agreement on its intrinsic value going forward. Putting aside vexing issues around finding an economic market for monographs, there was some agreement around the idea of the long form, however it is named, as offering unique value in the online environment. As our attention becomes increasingly fragmented, the book itself may provide even more service with its boundedness, so different from the open Web. It is contained; and decisions about what is inside and outside are quality and value decisions. As Richard Brown suggested, “The book is the anti-open-Web.”

A Note on Fair Use
No discussion of genre formation can elide the problem posed by the perceived failure of fair use exemptions to protect scholarly quotation and reproduction.
Fair use is essentially the right to use a small part of something under copyright for specific purpose. How that plays out online is controverted. A number of recent studies on fair use can shed some light on various media.¹ But we lack appropriate information to make sound judgments about what is and is not under copyright, such as databases in which to look up what rights have been retained for which purposes. There are some institutions with rich cultural materials that are ready to open access for broad use, such as Yale University, and some museums that are also favorable to sharing their resources. It is important to address an umbrella organization such the American Association of Museum Directors to push for freer access to museum collections.

Museum collections are encumbered by the considerations both of living artists and by potential donors. What will make a difference is providing alternative business models that allow museums and publishers to operate with freer access by offering in exchange such goods as capturing e-mail addresses, hosting community discourses to expand reach, and connecting their publics with scholars.

**Process, not Product**

As an increasing number of inventive models push out shared notions of production and authoring, there are some notes of caution to keep in mind. It is important at the beginning of each project to think about the lifecycle of the outcomes. We should not make “preserve it” the default mode for everything. This is where integration of projects with libraries as partners becomes crucial. A second note is to ensure that the product, no matter how open or closed, becomes part of a knowledge network. This will improve the project by inviting feedback. And it builds audience. But in a larger sense, it has evolved into a responsibility within the scholarly community to transmit experiential learning about how knowledge is created online.

What is the locus of such a network? Anne Helmreich cited a recent example of an art curator who asked about digital curation: where is he going to find out about how to do it? Where is the template for a digital catalog? How would he find a knowledge network to plug into? Establishing such readily accessible networks of knowledge that are able to keep current with best practices as they change is an important piece of digital infrastructure.

Julia Flanders warned against a slippage of terms that occurs as we go deeper into uncharted territory—for example, the frequent if unintentional muddling of digital, quantitative, and technological. These are distinct categories; reinforcing their distinctions in our debates will help give us some stream of clarity in what is a necessarily turbid river of discourses. Neil Fraistat, perhaps harkening back to the time of his specialty, the Romantic Period, before the word “scientist” was coined, urged us to lose humanities as a term, to the extent that it has come to connote “other than science.” Both appeals come from experts long adept at using digital technologies in the service of humanities scholarship, and indicate at a minimum that a time of deep engagement with new epistemologies and subtle, sophisticated reasoning around disciplinary formation is close at hand.

ASSESSMENT & CREDIT

The fundamental question—how do we identify and reward good scholarship—remains constant in humanities in the digital age. That said, assessment and credit for new-model scholarship cannot easily be picked apart from its creation and sharing. Nor is it a simple matter to pick apart audiences when we talk about assessment by whom and credit for whom. They beg new thinking, and devising new, appropriate protocols requires thinking by doing.

What to credit

There are numerous types of scholarship that go into the creation, sharing, curation, and stewardship of humanities content. Yet as rule, only monographs and articles—two specific types of final outcomes—garner their single authors credit and reward. (A list of numerous scholarly products was generated during SCI 8, and Cohen reproduced this list in his presentation at SCI 9.) Scholarly societies and their umbrella organization such as ACLS should begin conversations about micro-crediting—granting credit for different granularities of scholarly contribution, from review work to editing. Some argued that scholarly value extends far beyond creating knowledge; just as important and credit-worthy is being “a node of knowledge,” of discerning value, aggregating and curating content. They exert noteworthy impact on the direction of scholarly conversations. Different members of scholarly communication professions, such as scholars, librarians, publishers, programmers, designers, and others should receive reward and credit appropriate to their profession. Knowledge of the criteria for assessment across professional boundaries should be widely shared among cognate professions.

Who decides

Who constitutes the appropriate peer group? The argument for open peer review is that it broadens the pool of opinion to create a larger sampling and a smaller error rate. We are well aware that there are traps with impact factor and manipulating quantitative data. But there are also traps with small anonymous review pools as well, especially as the practices of scholarship are
metamorphosing so rapidly. Finding the individuals capable of reviewing new-model scholarship is not easy. Another advantage to open review is that it attracts an audience for content—or it has the potential to. The matter of finding time to review carefully does not change just because the review is open. Article-level metrics that measure downloads, time on page, citations, formal language analysis, experimenting with use of MESUR, Mendele, and Zotero should all be tested.

**LEADERSHIP & ORGANIZATION**

**Leadership Strategies**

Humanities is Janus-faced: studying and conserving millennia of human experience, at the same time focused on an unfolding future. At this juncture, when the core communication strategies of humanities are under such fundamental metamorphosis and ongoing negotiation, leadership among each of the professions of scholarly communication is critical. As mentioned, leadership takes the form of modeling and testing new practices, exploring foundational epistemologies and methodologies, and historicizing the present both to locate meanings and to bring the full human experience forward into the present. Peer-to-peer review and adoption has long been the standard of humanities ethics in practice; social and economic changes put a premium today on creating knowledge and sharing it.

The “build it and they will come” model of constructing digital infrastructure has proven notably inefficient and ineffective. A better strategy is to locate where people and energy congregate—both online and face-to-face—and add value. The approach then is “Ask not what the online world can do for the humanities; ask what the humanities can do for the online world.” Because the system of scholarly production and communication is a prestige economy, it is important to recruit high-impact individuals and prestige organizations to model practices and behaviors, publicizing, normalizing, and thereby making adoption of them desirable. Moreover, it is easy to overweight technology. Participants urged focusing energies around pressing scholarly issues, key intellectual and organizational questions of moment, and not designing the generic, the modish, the clever. Model, do not exhort. Build a community of discourse, not just your own CV.

What follows is a summary of the key points made about the professional sectors vital to scholarly communication; despite our rhetorical segmentation by familiar professions, we emphasize that the roles and responsibilities of contemporary professionals are extraordinarily fluid, with individuals and organizations playing multiple roles simultaneously.

**Organizations**

*Scholarly Societies.* Societies have the potential to become online nodes of deep knowledge. Amidst wide recognition that scholarly societies must reconceive
their member services for the digital age—given both the economic changes in revenue models tied to annual meetings and hard-copy journal publication and changing audience expectations—participants focused on the exact nature of the services they can provide to members and the larger public. The intellectual capital and prestige of scholarly societies makes them uniquely positioned to model forms of peer review of new objects of knowledge and new methods of review, as mentioned above. Individual societies are best suited to advocate for new standards of tenure and promotion that include digital work. Societies must show by example, not simply advocate: they can begin by featuring new-model scholarly communication processes and products in their journals, in their meetings, and in Web-based conversations they host. They are well-positioned to contribute scholarly authority to parts of the online world that need it, such as Wikipedia, or to add value to the digitized corpora found in Hathi Trust and Google Books collections. They can also develop new forms of bibliographies, citation protocols, and so forth. ACLS member societies can begin sharing their online resources amongst themselves, experimenting in nonmonetary “barter trades” such as reciprocal access to online resources. They should begin researching the new forms of scholarly network analysis, online commentary forums, and sorting through all facets of knowledge creation, curation, and sharing that warrant microcredits.

**Humanities centers:** The alliance between the [Consortium for Humanities Centers and Institutes](https://www.chci.org) (CHCI) and [centerNet](https://www.centernet.org) (the consortium for digital humanities centers) that resulted from conversations sponsored by SCI, is clearing a path to integrate the agendas of so-called traditional humanities centers and those focusing on so-called digital humanities. They are identifying numerous practical services they can provide to each other by sharing their expertise and methodologies, practices, and skills programs. More significantly, they are undertaking sustained investigation into research and programmatic agendas that become possible only by bringing the two groups together. Their initial areas of focus are [digital disciplines](https://www.chci.org/digital-disciplines), and [digital publics](https://www.chci.org/digital-publics). The first focus addresses disciplinary transformation emerging from new information technologies. The second focus is looking at the dynamic relationship between academic expertise and networked public knowledge. Both programs will result in deeper sharing and collaboration, eroding boundaries between theory and practice, providing a venue for assessing and re-crafting credentialing criteria, providing an unprecedented international reach, and developing what they style “the consortial imagination.” We know consortial action and collaboration is a precondition for success in the digital age. And we know it is difficult to engender and sustain collaboration. Therefore, this development promises to be uniquely significant for the humanities in and for the digital age.

**Libraries and publishers:** Bethany Nowviskie and Shana Kimball presented compelling examples of how libraries are providing new services for research and publishing in the digital age. The [Scholars’ Lab](https://scholarslab.lib.virginia.edu/) at the University of Virginia Library has rapidly emerged as a leader among library-based digital humanities centers by determining that, in addition to serving faculty needs, it is important to capture the attention and energy of technologists and graduate students. The Lab provides staff with dedicated time to pursue their own scholarly research
agendas and graduate students with the tools, space, and intellectual 
stimulation to stretch their imaginations, learn to collaborate, ask new 
humanistic questions of their materials and methods, and get early exposure to 
software development and design techniques. What has made the Scholars’ Lab 
a model for other universities is: first, that it is embedded within the library 
with rich primary and secondary sources and staffed with expertise across 
many academic disciplines and technologies; second, through fellowships and 
paid apprenticeships it provides students with opportunities for research and 
experimentation not otherwise available; and third, that it focuses as much on 
building communities of practice and networks of knowledge as building finite 
or fixed digital objects.

Kimball described the University of Michigan library’s grand experiment in 
bringing together the library, scholarly communication office, university press, 
and electronic publishing unit into one universe. The library is building on its 
traditional strengths to aggregate the many stages of scholarly communication 
and production, from the holdings in the library through its republication, use, 
further curation, and preservation. They are building and testing a new form of 
sustainability for publishing and scholarly production that embraces expertise 
in copyright, text creation, digital and analog preservation, and a venerable 
university press. Acknowledging that there is lots of overlap among functions, 
they are hoping that the overlap will actually provide an especially strong, 
responsive, and responsible infrastructure over time.

These examples of libraries expanding the scope of their responsibility in the 
digital age are not meant to be either proscriptive or prescriptive; but to provide 
reliable real-world models of how organizations deeply embedded within an 
existing analog infrastructure are rethinking the roles and responsibilities they 
have for online scholarship. Libraries, museums, and publishers all are facing 
extremely vexing choices. They are expected to maintain traditional services at 
the same time as moving swiftly into the future, most of which is only partially 
discernible. None of these organizations are well capitalized; finding the space, 
time, and resources to experiment and to risk failure for the sake of learning is 
perhaps the biggest challenge of all.

For some it will seem unnatural to lump libraries and publishers together. But 
the bifurcation of these complementary functions that occurred organically 
during the age of print turns out to put digital scholarship at high risk of 
corruption and long-term loss. University presses themselves are calling for 
closer working relationships with libraries, and the recent self-reflective report 
by AAUP acknowledges the need for fundamental change and imaginative 
alliances, beginning with reaching out to each other and to sister organizations 
on campus such as libraries.

**SHARED CAPACITIES**

The greatest common need identified by publishers has been to retool workflow 
for digital production. Streamlining workflow is inefficient and quite possibly
even impossible to do for each and every university press singly, on its own.
Optimal workflows do not demand one size to fit all, but they do demand a
different ecology of collaboration, beginning with identifying which aspects of
publishing and long-term curation and stewardship are best done in a
centralized fashion, and which are best left to local and disciplinary
customization. The consortial imagination is necessary for all sectors of new-
model scholarly communication, from scholarly societies to libraries, archives,
museums, and publishers. As scholars discover that doing scholarship online is
as much about building and sustaining communities of discourse as it is about
producing individual pieces of scholarship, we can imagine organizations
coming to a similar realization: that to work effectively in the digital
environment means to collaborate where economies of scale are critical; and to
grasp that there is more than enough room for multiple organizations to deliver
specialized products and platforms for project teams, disciplines, and
organizations.

The need for robust infrastructure across disciplines and campuses becomes
obvious when thinking about the demand of digital information for what is
known as lifecycle management, an integrated approach to persistent access to
knowledge by ensuring that from the time of its inception digital information is
created in formats that are technically, economically, and legally sustainable;
and that creators and curators understand distinctions among content
designed for obsolescence and designed to be sustained over time. University of
Virginia library director Karin Wittenborg reminded us that without digital
preservation, none of these efforts will be worth much in the long run or even in
the short run. Not everything created deserves to be preserved; some of it is
created for short-term purposes. Libraries need to keep decisions about
preservation relatively simple merely in order to deal with the volume of content
that is pouring in. What is optimal, she argued, is for scholars to do “self-
deposit” into a repository with specified retention periods. Effecting this change
requires scholars embrace knowledge curation as intrinsic to knowledge
creation. This is one aspect of developing an online persona that Scheinfeldt
mentioned—the capacity to make judgments about what should be sustained,
by whom, and for how long.

Both Don Waters of the Mellon Foundation and Josh Greenberg of the Sloan
Foundation called our attention to the importance of thinking creatively about
new divisions of labor and new models for collaboration, community, and
consortia. Each consortium will have its individual reason for being. Just as the
collaboration between CHCI and centerNet has well-defined goals to forge new
research and program agendas; and collaboration among several different units
at the University of Michigan is designed to build strong infrastructure for the
full lifecycle of scholarly communication, so we can imagine a series of
collaborations with discrete goals in mind, each important in the ecology of
new-model scholarly communication. What makes the ANVC of special interest
here is the ambition of visual culture scholars to develop new workflows for
collaboration among scholars, scholarly societies, libraries, archives, museums,
and publishing houses. This discipline-specific model may find very fertile
ground within scholarly communities in coming years. This approach also
provides opportunities for presses to come together and sort through which areas of digital scholarship they identify as their niche going forward.

Expanding a network of resources and sharing institutional capacities are important across the board, for no institution, no matter how well resourced, can build a standalone digital infrastructure. But beyond that, building shared infrastructure may be the only way to address the lamentable divide between digital haves and have-nots, a divide that is growing greater each year, separating public and private universities, comprehensive universities and liberal arts colleges, and various regions of the country. Although equal access to funds across the system is important, Josh Greenberg noted that there is already enough money in the system to move humanities forward. What hinders us is a combination of insufficient commitment to the well-being of higher education as an integrated system, and immature or inappropriate models—mental maps, if you will—of robust consortial infrastructure.

EDUCATION & PROFESSIONALIZATION

Education and professionalization are topics that were woven into every aspect of our discussions; it is misleading to pick them out from the context in which skills and expertise are needed. That said, scholarly communication professionals urgently demand changes in their preparation and opportunities for advancing on the job. The changes in roles and responsibilities taken on by organizations are occurring simultaneously with the emergence of new practices and expertise needed by these organizations. The skills necessary to be a producer and steward in digital scholarly production are significantly different from those embedded in the print model, and the differences go far beyond technical and computing skills. Most significant is, rather, a new way of thinking about how to identify and solve problems. Seeking the perfect solution, just like seeking a final, fixed version of scholarly argument, is counterproductive in a digital production environment. Change and evolution is the norm, and thinking coupled with experimenting and learning—the iterative process—is an appropriate and surprisingly efficient way to make progress both in solving problems and, in turn, properly identifying new opportunities and new problems.

Graduate students: On the majority of U. S. campuses, the integration of new technologies and practices into graduate education is ad hoc at best, and bordering on the negligent in some disciplines and on some campuses. Negligent in the sense that graduate students are unfairly and unrealistically expected to command digital literacies as a matter of routine, at the same time that they are told that these literacies cannot count for promotion or even in completion of the dissertation satisfactorily. There is widespread agreement that it is not until new graduate curricula which include core digital literacies have been developed, tested, and widely implemented that appropriate integration of 21st-century research skills will occur. Among the literacies identified as basic are:

- text mining
elementary programming
visualization
quantitative methods such as statistics
geospatial analytical and presentation skills
knowledge of intellectual property
project management skills
grant preparation
university administration
public writing and speaking

In addition to these basic skills imparted during graduate education, ongoing educational needs suggest short-term institutes that provide discipline-specific skills training and updating, such as those offered routinely by the medical and legal professions; some of these sessions could offer certificates. Institutes along the lines of the Digital Humanities Summer Institute should expand in number and scope, to be held both regionally and online. They can be convened and sponsored by scholarly societies. The latter could offer discounted rates for these training opportunities as a service to members.

New professions are emerging in scholarly communications, and the group that style themselves as alternative academics are forging a variety of new career paths as they pioneer the new modes of production and authorship. (Their work is documented in the #Alt-Academy project published by MediaCommons.) Many humanists with graduate degrees are veering off the straight and narrow path of tenure track to pursue their research and service agendas in entirely new ways. By necessity working collaboratively and as true entrepreneurs, they contribute to scholarship and to new organizational models. It is important they identify the working conditions they need to continue their pioneering work, for many of them face a clash between their expectations as scholars and researchers who can set their own agendas and claim ownership or control of their work, with the requirements of working in organizations key to infrastructure, such as libraries and presses, where the collective and managed enterprise works on different principles than that of individual scholarly careers. Among the questions this cohort points us to are the ways undergraduate as well as graduate education should be changing to prepare students for advanced literacies in the digital era. They also demand attention to the purposes of the dissertation as professional training ground when the profession at the end of training is an alternative to the traditional tenure-track.

SCI participants called on organizations such as CLIR with its longstanding postdoc program for humanities PhD, to work through the vector of the #Alt-Academy project to survey alternative academics and their employers for perceived gaps in professional preparation.

**FUNDING & SUPPORT**

The attendance of numerous funders at SCI is but one indicator of the changing landscape of humanities funding. The current precipitous decline in public-
sector support for higher education hits the humanities with particular force in undergraduate and graduate schools. Simultaneously, there are an increasing number of private-sector funders beginning to recognize just how fundamental humanities education is in the digital age. The challenge for them is to identify points of strategic intervention in a rapidly changing landscape that would either incubate or accelerate desirable change. Several participants noted that the collapse of the remnants of the Cold War public-sector funding strategy, with its emphasis on instrumental educational means towards instrumental social and political ends, has created gaping holes in knowledge of foreign languages and cultures precisely the moment when they are in greatest demand. The need to find new funding sources for these and essentially all humanistic competencies means that we need new funding strategies. Which funding streams will create digital infrastructure, encourage focus on scholarship itself, and, at least in the short term, produce a number of “problem factories” that challenge received wisdom about the best way to do things and point us in new directions?

One of the new directions participants returned to time and again is the need for building mutual dependencies to create economies of scale and strengthen ties among the sectors of humanities who share values and goals. Mobilizing communities for "digital philanthropy" will require making a compelling case for their engagement, and helping to identify strategic interventions has become the urgent work of scholarly societies, libraries, publishers, and all leaders in scholarly communication. Participants brainstormed the needs and opportunities that need outside support. They include:

- Supporting experiments for collaborative work among libraries, presses, scholars
- Convening conversations to build knowledge networks
- Incubating new organizational and consortial models
- Building and donating digital collections, software, hardware, dedicated laboratory space, and so forth

**VIEWS ONTO THE FUTURE**

Trying to imagine an ideal system of scholarly communication without falling into the trap either of projecting past models into the future or lapsing into technological determinism has been an ongoing challenge throughout the history of the Scholarly Communication Institute. We have been fortunate to have participants well grounded in deep and deeply historical humanistic thought able to offer guidance. At SCI 2, which addressed the emerging discipline of practical ethics, the philosopher William May proposed that what faces us is essentially a moral challenge, one that the humanities has faced often. The vector of communication in the Academy settled long ago on the vertical, from the mentor to the mentored and back. Digital technology favors the horizontal over vertical communication; this offers humanists a longed-for opportunity to communicate both with each other and with the interested public. Humanists have a fine and nuanced knowledge of the human condition
in historical time, and we have an obligation to share our knowledge clearly, directly, non-dogmatically. In 2004, May predicted that this would lead to necessary changes in the genres favored by humanists, moving us back toward its originating form of the humanistic essay. Whether our new technologies result in a revival of the essay or the emergence of a novel genre, they do prompt us to re-examine and re-engage the fundamental means and ends of the humanities.

As debate at SCI inevitably gravitates towards discussion of review and credit, the questions of the norms of humanities scholarship have been raised consistently and with some consternation. What is humanities scholarship, and what is humanities research? At SCI 8 (2010) David Brownlee, an architectural historian and editor of the Journal of The Society of Architectural Historians, reminded us that humanities scholarship is what humanities scholars do. In midst of often difficult discussions about credentialing it is wise to remember that the desired outcome is not a perfect replication of the system and its tenure genres of monograph and article. It is the conscientious stewardship of human knowledge over time for the benefit of past, present, and future generations. That would argue today for the current generation of senior scholars to put in place for their successors a system that allows the same scope of freedom to define individual research agendas as they inherited from their predecessors. In the end, what we are striving for is a scholarship in which the questions of audience, attention, and authority in the online world remain provocative, discomfiting, difficult to resolve, and open for debate.

We asked a group of scholarly communication professionals who are building the humanities in the online world to describe an ideal environment that would nurture and support their ambitions, keeping in mind that we need to demonstrate value not just to review committees, or to each other, but to the public. They homed in on a set of principles, features, and actions they advised people to foreground as they develop an agenda for further exploration and experimentation.

In the 21st century, the production and communication of knowledge are processes that are inherently dynamic, interrogative, and dialogic. The model of the fixed expression of knowledge can be seen as a historical artifact of the need to fix an expression onto a durable form (stone, paper, film) to ensure ongoing access. The digital, however, returns us to a state of plasticity, similar to that of oral culture, that demands managing the lifecycle of knowledge in a dynamic and flexible infrastructure. The process of moving humanities into the online world will not be accomplished in a short period of time. Neither, in the long run, can we expect to see a system that is perfected and static. Nor will the scholarly communication professions be siloed into those who create, those who catalog and curate, those who preserve and serve, those who publish, those who administer, and so forth.

In such an environment, the necessary conditions for digital production and communication suggest building infrastructure along the following lines:
• *reduce risks of experimentation*—encourage what Kimball calls "strange institutions" and hybrid forms that would build on deep disciplinary knowledge and scholarly grounding, yet set aside the inherent conservatism of humanities disciplines and professions

• *connect and expand*—consortial thinking and efforts can spread and dilute the risks of experimentation, address scale and resource needs, and work to establish communities of practice

• *lower unnatural barriers*—redress the inadequate education and preparation for the new work of humanities; revise outmoded information and intellectual property policies; retool systems of recognition and reward

• *collaborate across unnatural boundaries*—libraries, presses, administrators, designers, and programmers should work with, not for scholars; collaborators must establish clear rules of engagement that respect natural divisions of labor, not those aligned with outdated and empty distinctions of prestige

• *acculturate and include*—nurture communities of discourse that can model behavior, socialize new forms, encourage learning by doing, and expand audiences

• *model behaviors*—embrace an iterative process, both for the purpose of learning by doing and for the purpose of modeling and socializing new forms

**WHERE ARE THE NEW RESEARCH QUESTIONS?**

Michael Steinberg noted that first-wave use of technologies have by and large supported existing methodologies and questions. Miriam Posner concurred that "most research using new tools seems to pose not radically new questions, but different versions of the same questions. The biggest change I've seen is in the tools that people use, not necessarily their methods." There is a tendency to confuse a new mode of argumentation—such as the use of multimedia—with a genuinely new research question. Part of this may be due to the fact that some funding streams have specified tool development as a goal in and of itself. Part of it is due to the inherent conservatism of disciplines, which increasingly differentiate themselves not by the subjects they address, but rather by the different methodologies they use, how they constitute a problem, and how they recognize and reward achievement.

And yet there has been an expectation for well over a decade that the radical refashioning of information technologies from analog to digital will change how scholars writ large identify, solve, and generate problems. We have witnessed new topics of research emerge—the environment, gender and the body—across
several different fields. But they have seldom been tackled anew, outside the usual sociologically distinct disciplines embedded in U. S. universities.

Technologies matter in disciplinary formation and development of new research questions to the extent that communities of practice arise around specific technologies, and that leads to communities of discourse—the origins of disciplinary alignments today. We know that new technologies allow for better pattern recognition, a long-standing interest both for humanists and scientists seeking both what is normative and what is exceptional in human and natural worlds. And we see initial forays into questions that bring several disciplines together, such as medical humanities bringing together the neurosciences, clinical sciences, performing arts, ethics, and narratology, among others. We see emerging interests that call together neuroscience, acoustics, musicology, sound studies, and other fields to study the aural ecologies we inhabit. Moving into new fields entails as much attention to disciplinary cultures as well as epistemologies, and this is an area where the potential for face-to-face discussions that identify both leading topics and leading scholars—cross-sectoral meetings such as SCI holds—could prove fruitful.

SCI was designed to foster experimentation, articulate the needs of online scholarship and the infrastructure to support it, and then get out of the way. Scholarly production and communication increasingly incorporate the cycle of imagination, experimentation, reflection, and further imaginative provocation. Since the inception of SCI, participants have noted a historic shift from thinking of the ends of scholarship as process rather than product. What has become yet more clear over the decade of SCI is how not only the end product of scholarship is in motion, but the scholarly communication system itself is “in process,” infused with a dynamism that was unthinkable 20 years ago. As the vector of communication expands horizontally ever outwards, the humanities’ core enterprise is to join in and lead the larger social project of fundamentally remaking audience, attention, and authority.

*****

NEXT STEPS

Scholarly Production & Authoring:

Actions:
1. Continue and expand ongoing experiments in new platforms for publishing and new genres (Scalar, Press Forward, Media Commons and MLA’s Digital First).
2. Document and compile results of these experiments.
3. Develop metrics for engaged use, influence, impact, and lasting value to scholarly discourses.
4. Incorporate mechanisms in production and authorship that ensure projects can be engaged at different points of development.
5. Build new business models that support nested granularity.
7. Address fair use concerns by establishing disciplinary bodies of practice, extending reciprocity agreements among academic publishers, working with museums, libraries and archives to expand access to content and develop alternative business models to compensate for lost revenue streams.

**Actors:** scholarly societies, humanities centers, publishers, libraries, funders

**Assessment & Credit:**

**Actions:**
1. Articulate benchmarks of scholarly merit in digital scholarship
2. Undertake scholarly network analysis.
3. Develop readily adoptable peer review and commenting systems for post-publication assessment.
4. Develop and publicize assessments and micro-crediting systems for all professionals in scholarly communication

**Actors:** scholarly societies, ACLS, humanities centers, libraries, publishers, administrators

**Shared Infrastructure:**

**Actions:**
1. Develop partnerships among scholars, libraries, and publishers to support new, streamlined production and use workflows that operate throughout the lifecycle of a digital creation.
2. Develop cross-institutional collaborations between presses, libraries, research centers.
3. Extend reach of digital laboratory environments.
4. Build networks of knowledge and explore mechanisms for continuing cross-sectoral conversations to share knowledge and accelerate collaboration.

**Actors:** scholarly societies, ACLS, CLIR, humanities centers, libraries, publishers

**Education & Professionalization:**

**Actions:**
1. Develop new curricula and apprenticeship opportunities that address the actual skills required in scholarly communication professions.
2. Develop new curricula for research methods, project management, design and editing skills, public writing and speaking.
3. Increase numbers of and access to venues that provide lifelong education in new-model scholarly communication, such as regional institutes, distance learning courses, and so forth.

4. Investigate the appropriateness of the dissertation as now practiced for preparing graduate students either for a lifetime of sustained scholarly productivity or for other intellectual but nonacademic career paths.

5. Develop a network of labs and centers in which graduate fellowship and practicum programs can be tested and best practices shared.

**Actors**: scholarly societies, humanities centers, publishers, libraries, funders, SCI

**Funding & Support**:

**Actions**:

1. Engage new sources of support for digital collection building, professional development, library and curatorial skills.
2. Articulate a compelling case for humanities in and for the digital age by documenting, aggregating, synthesizing, and publicizing concrete contributions.
3. Incubate and accelerate new models of consortial thinking.
4. Provide startup funds to experiment with new models for peer groups, university presses.

**Actors**: SCI, ACLS, CLIR, private-sector funders
No Brief Candle: Reconceiving Research Libraries for the 21st Century

August 2008

Council on Library and Information Resources
Washington, D.C.
Contents

Foreword ................................................................. v
Conference Participants ............................................ vi

Part I: A Continuing Discussion on Research Libraries in the 21st Century . . . 1
Introduction ............................................................ 1
Prologue to a Fundamental Rethinking: Context and Topic Threads ............... 2
Reconceiving the Library: A Vision for the 21st Century ............................. 8
Challenges and Constraints ........................................... 9
Recommendations ....................................................... 10

Part II: Essays ........................................................ 13
The Research Library in the 21st Century: Collecting, Preserving, and Making Accessible Resources for Scholarship, by Abby Smith ............ 13
The Future of the Library in the Research University, by Paul N. Courant .... 21
Co-teaching: The Library and Me, by Stephen G. Nichols ......................... 28
The Role of the Library in 21st Century Scholarly Publishing,
    by Kate Wittenberg .................................................. 35
A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries, by Richard E. Luce ......................... 42
Accelerating Learning and Discovery: Refining the Role of Academic Librarians, by Andrew Dillon ......................................................... 51
Groundskeepers, Gatekeepers, and Guides: How to Change Faculty Perceptions of Librarians and Ensure the Future of the Research Library,
    by Daphnée Rentfrow ................................................. 58
Leveraging Digital Technologies in Service to Culture and Society: The Role of Libraries as Collaborators, by Lee L. Zia ......................... 66
According to Samuel Johnson, “Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information upon it.” Until recently, we knew where we could find information upon any given subject—in a research library. Libraries collected, preserved, and made available an array of resources needed by scholars. The bigger and more comprehensive the research library, the greater was the community’s access to knowledge, as well as access to those experts who could help patrons navigate the library’s geography of knowledge. Because scholarship has been primarily print and artifact based, the library was bound to acquire and then maintain in usable form scholarly literature and primary resources in order to make them accessible. In hindsight, it seems unlikely that between them, so many libraries would have redundantly purchased so much of the non-unique secondary scholarly literature if they could have made it accessible to their patrons in less expensive ways—ways that did not demand large and continuing investments in physical, technical, and staff infrastructure. The success of interlibrary loan gives some evidence to this surmise.

Whereas libraries once seemed like the best answer to the question “Where do I find...?” the search engine now rules. Researchers—be they senior scholars or freshmen—no longer make the library the first stop in their search for knowledge. The shift from producing and consuming information in hard copy to multimedia digital form has moved the center of information gravity from research libraries to the Internet, and done so in a dramatically brief period. The preconditions for this sudden shift were laid in the 19th century by the development of audio and visual formats—still and moving images, recorded sound, and, ultimately, formats combining...
sound and image. A bifurcation eventually emerged between campus-based “general collection” libraries, which focused on secondary literature and a highly selective group of primary sources (both print and nonprint), and libraries not serving first and foremost a faculty and student body, and which focused on “special collections.”

I mention this division of labor among research libraries because it is a mistake to grant exclusive agency to digital information in the shift away from the centrality of academic research libraries in collecting and preserving resources for scholarship. The academic research library has been predominant in collecting and preserving text-based scholarly literature, but it has not been the primary home for statistical data, cartographic materials, manuscript collections, prints and photographs, film, broadcast television and radio, folklore documentation, natural history specimens, and an overwhelming preponderance of primary source materials needed by scholars in the humanities, social sciences, and physical sciences. The challenges facing academic research libraries are fundamentally different from those facing nonacademic research libraries, not because of their mission (they both serve scholarship) but because of their user base. I will focus my remarks on the former because they are facing more urgent pressures to change, and so emerging trends for research libraries of all stripes may be easier to comprehend.

If we take libraries-as-first-resort in search out of the equation, what is left looks something like stewardship, loosely defined: ensuring long-term access to content in reliable, secure, and authentic form. But we already know that a significant portion of digital scholarly literature and primary resources—that is, the portion available through licensed agreements—is seldom in the possession and care of research libraries. Perhaps a preliminary answer to the question “What are the core functions of the research library with respect to collecting, preserving, and making accessible resources for scholarship?” might be that research libraries will be stewards of some sectors of the information universe, but they will not be the same sectors as before. So which sectors will they be?

Collecting, Preserving, Making Accessible: Where We Are Headed

To answer that question, we will examine six trends in the academic research environment that are likely to shape scholarship in the next decades. From these trends we may learn something about what resources scholars will use and how. First, however, I believe that one thing about scholarship will never change: scholars will demand access to information resources to examine what others have discovered and thought; to use and reuse evidence and scientific conclusions; and to publish results of their own research based on these resources. That is why their sources must be authentic, reliable, easy to find and retrieve, and easy to use and reuse.

1 Separately incorporated “special format” libraries on campus share features with both types of libraries; how much varies greatly depending on how closely each is integrated with and funded by the main university library.
1. Ascendance of science

The physical and life sciences are expanding their footprint on almost every Research I university campus. Science programs have become huge cost centers, consuming an ever-larger portion of university expenditures on research infrastructure. Because of the way science is funded, however, these programs are often viewed on campus as revenue centers: they are recipients of the largest federal grants and the largest philanthropic donations, in the tens and hundreds of millions of dollars. Science is where the big donors like to make their mark, comparable to the way that libraries were magnets for philanthropic donations in the 19th and 20th centuries. So science, which costs universities a great deal, will nonetheless increase in charisma; and the humanities, which neither cost so much nor bring in so much money, appear at present to be politically disadvantaged.

But that is just the money side of the equation. More significant in the long run is the influence of scientific reasoning on nonscientific domains of research. There is a general expansion of quantitative reasoning and methods into normally qualitative disciplines. For example, imaginative uses of geographic information systems (GIS) in history, archaeology, and art history, and data mining in classics and other text-driven disciplines are breathing new life into old disciplines. There is a burgeoning demand among social scientists to incorporate into their research an array of scientific data—such as epidemiological information and distribution patterns of genetic variations with health care statistics—and methods, such as GIS-based geographical analysis used to plot and examine polling or census data, consumption patterns, and so forth.

Finally, more and more scientists are recognizing that persistent data management is crucial to their research. Hence, they are developing library-like centers for the collection, curation, preservation, and access of data. The National Science Foundation has encouraged them to do so by putting out a call to develop such structures as key components of scientific cyberinfrastructure. Private foundations, including the Alfred P. Sloan and the Gordon and Betty Moore Foundations, are making equally significant investments in scholarly communication that include stewardship as well as dissemination.

2. Development of digital humanities

The accelerated development of digital humanities is an even more significant trend for research libraries, if only because humanists have been their primary clientele. Beyond the increasing use of quantitative research methods in the humanities, there is a growing demand by humanists to access and manipulate resources in digital form. With the primacy of “data-driven humanities,” certain humanities disciplines will eventually grow their own domain-specific information specialists. While perhaps trained as librarians or archivists, such specialists will work embedded in a department or disciplinary research center.
Of greater import is the emergence of digital humanists who continue to focus on narrative, discursive, and essentially qualitative ways of investigating what it means to be human. It is these scholars, interrogating new forms of discourse, narrative, communication, community building, and social networking, who will spend most of their time on the open Web and use wiki and blogging applications, social software, and other as-yet-undreamt-of applications. All these multimedia forms of discourse will present special challenges for collection development and preservation because of their inherent bias toward process over product, a bias that resists fixing expression in the canonical forms upon which analog preservation practices are dependent.

3. Emphasis on process over product (with respect to scholarly communication)

Distinctions between formal, archival publication and informal modes of scholarly communication are becoming nebulous. Among scientists, we have seen for more than a decade a preference for various types of informal, preprint-type sharing of working drafts, an informal mode of communication that has greater impact on the development of scholarship than the final, archival or formal publication does. (The latter, however, will probably continue to have a greater impact on scholarly careers, at least for the short term.) Humanists are also becoming more engaged with informal, narrative forms of communication, with graduate students and tenured professors alike using vernacular social software applications to build communities of discourse.

What does this mean for scholarly communication? I recently heard a tenured literary theorist say that she hoped never to publish a monograph again. When she gives talks, they are immediately blogged, and she finds this mode of discourse with other scholars highly productive and immediately gratifying. It has also reframed her view of the timetable of monograph production, shifting from inevitable-if-slow to arbitrary-and-obsolete. So much for the time-honored notion that humanists are immune to the pressure of time to get out their research results!

Finally, in many domains we see an erosion of the traditional distinctions between primary and secondary sources and flows of information. Many scholars now argue that publication and dissemination can and should represent evidence as well as argument, and that is precisely what they demand of new-model scholarly communication.

4. Mobile and ubiquitous computing

The headline here is that the laptop is the library. It was recently reported that a researcher at IBM is working on a storage technology that will allow an entire college library to be stored on mobile devic-
es as small as the current iPod. Whether it happens two years or five years hence, whether it is IBM or some other company that realizes this goal, the handheld library is foreordained. Even without such a device in hand, we see the dominance of consumer technologies and applications, both commercial and free, in the academy. It is not only the undergraduates who arrive on campus with iPods that can stream courseware and the senior faculty who consult just-in-time Web-based references, even offline, through Zotero. It is that undergraduates can have a sophisticated command of geospatial thinking simply by opening up Google Earth; they do not have to master the intricacies of GIS available through expensive ESRI applications. It also means that graduate students do not require a high-quality but expensive (and far from ubiquitous) resource like ARTstor for creating presentations, sharing links, and drafting articles, when an astounding number of equally high-quality images are available free on Flickr. Then the question for research libraries becomes how to provide persistent access to these sources. Or does it? Does that become someone else’s responsibility?

5. Data deluge
Given the scale of information that scholars must cope with daily, opportunities to acquire skills in information management should be a key element of their education and training. The goal of professional training as a scholar is to maximize the autonomy and enhance the creativity of the scholar as an arbiter of information. We should never underestimate how carefully successful scholars manage their time; ready access to information that fits within the time frames set by the scholar is often the most important criterion in information seeking. Only some aspects of scholarship demand information meeting the rarefied benchmarks of reliability, authenticity, and persistence. That is why many scholars begin searching for information on the Web, and why they often turn to Amazon.com, not their local OPAC, to do a “quick and dirty” literature search.

With one more stage of breakthrough in storage, we could see significant change in the way individuals are able to manage the data deluge. The device under development at IBM, mentioned previously, “could begin to replace flash memory in three to five years, scientists say. Not only would it allow every consumer to carry data equivalent to a college library on small portable devices, but a tenfold or hundredfold increase in memory would be disruptive enough to existing storage technologies that it would undoubtedly unleash the creativity of engineers who would develop totally new entertainment, communication and information products.”

3 Ibid.
6. Rising costs and changing funding models for higher education

Competition for funding among all units on campus means that the library must continuously demonstrate its value; it must also bring in money or lower costs simply to provide services demanded by their users. Given the financial pressures on all aspects of higher education, it is imperative to change the service model of the library. When the world was smaller, libraries strove to be many things to many different constituents. The library must now focus on specific communities. Its role in pedagogy seems clear, as pedagogy is always locally based. But an individual library’s role in research, an increasingly global enterprise, is not so clear. Each research library will need to find its niche. This is why the “special-collection” research libraries that have a tradition of being subject or format based may, in the long term, be better models for research libraries than campus-based general-collection libraries are.

Collecting, Preserving, Making Accessible: Two Roles for the Library

So what can we infer from these six trends for the research library with respect to scholarly resources? First, let us define the research library as a line item in a university budget dedicated to managing information resources for research and teaching. For our purposes, it matters little whether in 25 years that function will be performed by something with the discrete name of “library.” Whatever its name, that entity will need to focus clearly on two specific roles: one local, the other networked and part of a national and transnational research cyberinfrastructure.

In its local role, the library will be optimized to meet the needs of its campus community. The library is likely to provide repository infrastructure for stewardship of university-based information assets. Most of those assets will support pedagogy, administration, student life, alumni affairs, and other things vital to the school. A much smaller portion of them will support research. Research will be a far more global phenomenon than local institutions can support on their own.

In its networked role, the library will be able to support research and dissemination to the extent that it is tightly networked into the increasing cluster of inter-institutional collaborations that enable the creation and use of scholarly content. These collaborations will be key elements of research cyberinfrastructure, an infrastructure that will be a research-and-dissemination platform. In the magic phrase of the digital era, it “will scale,” be ubiquitous, and support a variety of scholarly domains, from astronomy to nanobiology, archaeology to urban design. The next-generation research library must be firmly

---

4 A characterization recently used by Kimberly Douglas, university librarian at Caltech, that distills the relationship of the library to a community of scholars and why it can command so much university money.
embedded in that infrastructure, because that will be the platform to which scholars will gain access on their laptop library.

The exact models of stewardship and dissemination in the cyber-infrastructure will be determined by the evolution of domain practices. In the quantitative fields, we see domain-specific stewardship models such as genome and protein databases, the Virtual Observatory, and the Inter-university Consortium for Political and Social Research (ICPSR), among others, that look quite similar to “special-collection” libraries writ large. These entities are scaled to collect, preserve, and make accessible digital research content. They are deeply embedded within the communities of researchers that they serve. These stewardship models are optimized to handle content created by and for the academic community.

These networked efforts should also be extended to the data that are created outside the dominion of the academy, of particular value to humanists. This content comes in roughly two flavors—commercially created (usually gated) and publicly networked (ungated). So far, one organization focusing on stewardship of publicly networked content—the Internet Archive—has achieved scale. It is so successful at this that it provides vital services for numerous national libraries and government organizations seeking to archive their domains. Scarcely a decade old, it is already indispensable. While scores of university research libraries are collecting Web-based content in selected areas, none of them achieves, or even aims for, the scale or breadth necessary to collect digital content that scholars will demand. While I believe that certain research libraries can achieve comparable scale in collecting, it is unclear that any are planning to do so, or that they even see this as part of their core mission. It is equally unclear which libraries, if any, other than the Library of Congress, are contemplating large-scale partnerships with commercial content providers to ensure long-term access to primary digital resources. This is bad news. In the absence of such efforts, researchers will be forced to rely on commercial entities to preserve and make accessible their own content on their own terms.

Where academic libraries have been more effective, not surprisingly, is in joining networked efforts, such as LOCKSS, CLOCKSS, and Portico, to ensure persistent access to scholarly literature. These are important efforts and have much to teach about the challenges of forging long-term trusted relationships that can ensure access to content over time. For this is the make-it-or-break-it challenge for academic and nonacademic research libraries alike: to forge close working relationships with content providers—be they individuals, for-profit corporations, or learned societies—to ensure persistent access to that content for generations to come.

**Conclusion**

Research libraries evolved over the course of centuries to solve the problem of providing access to information. The library was the place where the artifacts of knowledge were aggregated and individ-
uals came to consult them. The stewardship of artifacts will continue to be a collective responsibility of the research library community. As more of their content becomes available through digital surrogates, more opportunities will open for libraries to design a collective solution to preserving the artifacts.

But if we were to design a system to address the needs of digital scholarly resources, it would certainly be different from the library. The system would combine the functions of library, information technology, and scholarly publishing. Those who manage information resources for research and teaching would take it as ground truth that research is a global and distributed phenomenon. So, too, should be the infrastructure that undergirds it. These managers—be they called librarians or not—would be responsible for building and maintaining the multiple partnerships with scholars, learned societies, content creators, publishers, and, above all, with each other across the globe, that would support persistent access to high-quality research resources.
On the following pages, you’ll find a set of templates for roughing out new interface and user experience ideas for your RBS case study or real digital project. We recommend pen-and-ink or colored pencils, and a carefree attitude: the quality of your thought is much more important than the quality of your artwork!

Use the smaller templates to sketch many competing possibilities quickly—and, later, to tell a story about your website by showing a sequence of screens or interactions. Use the larger templates to explore a single screen or idea in more depth.

Above all, use these templates up! And then sketch or photocopy more, or find alternative templates for yourself online.

YouTravel SEARCH

- Get people to EnOugh
  Info and "feL" of
  a Trip to decide to
  commit + get more info.

- Trips are Visual:
  Maps, Photos, Places,
  Events.

- Time may Matter
  (Winter = Ski
   Fall = Scenic Drives)

- Sometimes Location
  is already decided
  or (Grand Can't You?),
  sometimes not so much
  (The Beach? a Get-Away?)

- What about $45??
You Travel Search

Your search for Tahoe found more than one location:
- Lake Tahoe, NV
- South Lake Tahoe
- Tahoe Ski Resorts

Browse by photos, maps, prices, or guides.

Trip Preview: Ski the Big 8!

Why 6?

This trip

Good for:
- Family
- Friends
- Couples
- Adventure

Flag it / Hide it

More

Note: This is a complete fake project!