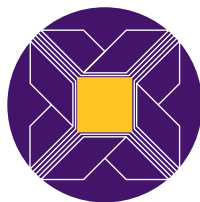


# The Research Library's Role in Digital Repository Services

Final Report of the ARL Digital Repository Issues Task Force

January 2009



Association of Research Libraries

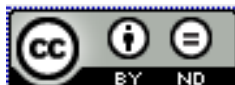
The Research Library's Role in Digital Repository Services  
Final Report of the ARL Digital Repository Issues Task Force

January 2009

Acknowledgements

The task force members are grateful to a wide variety of individuals who willingly shared their expertise with them. In particular, Clifford Lynch made singular contributions to the group's work and the final report, generously providing his singular and always insightful perspective. Wendy Gogol, Catherine Mitchell, Bernie Reilly, Tracey Robinson, and Mackenzie Smith provided information on their experiences leading digital repository projects and a wide range of helpful insights based on that work. Lars Meyer generously shared the conceptualization presented in the figure on page 12. Finally, the Task Force would like to acknowledge the contributions of several individuals who offered very substantive comments on various drafts of the report: the members of the ARL Scholarly Communication Steering Committee, ARL E-Science Working Group, Amy Friedlander, and Deanna Marcum.

Published by the  
Association of Research Libraries  
Washington, DC 20036  
[www.arl.org](http://www.arl.org)



This work is licensed under the Creative Commons Attribution-Non Commercial Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

# Contents

---

Preface .....	5
Executive Summary .....	7
Introduction and Framing of the Arena .....	11
<i>Questions and Frames</i> .....	13
Common Issues and Choices .....	15
<i>New content plus old content in new forms</i> .....	15
<i>Engaging with a larger networked digital environment</i> .....	17
<i>Attending to the "Demand Side"</i> .....	18
<i>Sustainability</i> .....	19
Strategies for Service Deployment .....	23
<i>In-house development and deployment</i> .....	24
<i>Collaborations and partnerships — within institutions, across institutions</i> .....	25
<i>Balancing local and vended services</i> .....	28
<i>The bigger picture</i> .....	28
A Horizon Analysis .....	31
1. <i>The user environment</i> .....	31
2. <i>The technology environment</i> .....	32
3. <i>The shape of library collections and services in 2015</i> .....	33
4. <i>The policy environment</i> .....	33
Positioning Research Libraries for New Roles .....	35
Call to Action to ARL Member Libraries .....	39
Conclusion .....	41
Selected Bibliography .....	43
Appendix .....	47
<i>Summary of responses from other organizations regarding activities around digital repositories</i> .....	47

## List of Figures and Sidebars

Figure. Elements of Digital Repository Services .....	12
The Example of MIT's CWSpace .....	16
The Example of arXiv .....	20
The Example of Harvard's Digital Repository Services .....	24
The Example of Scholars Portal .....	25
The Example of HathiTrust .....	26

## Preface

---

Early in 2008, the Association of Research Libraries (ARL) convened a Digital Repository Issues Task Force to evaluate trends, contextualize current activities among ARL member libraries and recommend leadership roles and activities. In doing so, ARL created a group of individuals with expertise in various domains from institutions currently engaged in digital repository development. On meeting for the first time to consider our own experience and the broader landscape, we were immediately struck by the diversity of types of digital content being managed by our institutions. Existing digital repository collections include:

- published faculty research archived for institutional purposes
- unpublished text material from faculty
- research data in various numeric and image formats
- administrative records
- primary source documents from libraries and research centers
- digitized book, journal and image collections
- instructional materials and courseware
- platforms for publishing journals
- software

We began our work by looking for common issues arising from this wide array of repository types and concluded that indeed, there are several concerns and questions that have emerged from the various pioneering efforts in repository building.

Very early in our analysis, we recognized that there has been a great deal of work done which addresses

technical issues related to the building and maintaining of digital repositories. That work is ongoing, and some of the major activities are noted in the appendix to our report. Our task force focused on the repository services that our libraries could provide for the research university community. A scan of major repositories currently in operation revealed a variety of strategies for offering services. We found their examples instructive and hope that others will also find the brief summaries of their experiences useful.

Before moving to recommendations for ARL member libraries, the task force considered the impact of trends in the broader environment. We attempted to forecast what we can expect by 2015 from four perspectives — library users, technology, library collections, and the policy environment. Looking ahead at the changed world in which libraries will operate in a very few years, we recognized that it is essential that research libraries act immediately to position themselves for new roles. We all need to develop a deeper understanding of content users' and creators' needs, and we need to develop services based on the life cycle of research information.

We urge ARL libraries to reflect on the predictions outlined in the Horizon Analysis section and how library services will be affected. Through this serious consideration of undeniable trends, we became convinced of the critical need for all research libraries to become engaged with digital repository development, an essential function in future library services.

Because of the urgency of the need, we have framed our recommendations as a call to action to ARL member libraries. Building appropriate partnerships within

and beyond local institutions is a vital starting point. We need to collaborate not only among ARL member libraries, but also with other key members of the larger networked research environment. Indeed, the diverse expertise represented by the members of the task force enabled us to see beyond our usual perspectives in addressing the questions facing us all. We must demonstrate a willingness to try new approaches, to evaluate early experiments, and to expect to make adjustments in order to develop the innovative services that will define our future roles.

In the period since our group was first convened, the economic environment in which our institutions operate has weakened considerably. The recommendations we make, however, are crucial for the longer term, regardless of the financial constraints we face in the short term. As we are forced to reconsider activities in the current downturn, we urge ARL members to look at digital repository services as a strategic priority for the future, and ARL to incorporate our findings into program planning.

As a task force, we thank ARL for the opportunity to learn from each other as we explored research libraries' opportunities to establish new roles through delivering digital repository services.

**Task Force Members:**

**Carole Moore**, Chief Librarian, University of Toronto  
(Chair)

**Karla Hahn**, Director, Office of Scholarly Communication,  
ARL

**Gregory Jackson**, Vice President and Chief Information  
Officer, University of Chicago

**Brian Lavoie**, Senior Research Scientist, OCLC

**Sarah Pritchard**, Charles Deering McCormick University  
Librarian, Northwestern University

**Oya Rieger**, Associate University Librarian for  
Information Technologies, Cornell University

**John Wilkin**, Associate University Librarian for Library  
Information Technology & Technical and Access Services,  
University of Michigan

**Ann Wolpert**, Director of Libraries, Massachusetts  
Institute of Technology

## Executive Summary

---

### *Introduction*

Digital repositories are developing rapidly as a key element of research cyberinfrastructure. Even when research institutions are grappling with difficult budget decisions in the current economic environment, they need to have a strategy for providing repository services.

Libraries are making diverse contributions to the development of many types of digital repositories, particularly those housing locally created digital content — including new digital objects or digitized versions of locally held works. In some instances, libraries are managing a repository and its related services entirely on their own, but often they are working closely with other stakeholders at their institutions to jointly develop repository services.

The Association of Research Libraries (ARL) Digital Repository Issues Task Force was charged “to evaluate trends, contextualize repository activities among ARL libraries, and recommend leadership roles and activities for ARL.” It embarked on its work at a time when many formative repositories exist, making it easier to begin to identify dynamics that will shape repository development. Institutional repositories are a common form of repository, but this report focuses more broadly on the full range of repositories. At the same time, it concentrates on repository services rather than repository technologies or content. Repository services include services to authors, contributors, and users, particularly of university-created content. Some examples of repository services provided by research libraries include long-term archiving and migration of content, dissemination and access management, metadata and format management,

search and discovery tools, publishing, data mining, etc. Illustrations drawn from a variety of digital repositories are used throughout the report, demonstrating the value of the wider perspective.

Repository services are built upon a foundation of content, context, and access, requiring a balance of investment and emphasis among these three elements. The early institutional repository movement emphasized access, and in many instances institutional repository service development has struggled to build the content element. On the other hand, repositories coming out of digitization programs tended to initially emphasize content and continue to struggle with context issues such as metadata creation and standards support. Jointly, context and access issues underlie many of the emerging challenges of supporting discovery services. Collectively, all three issues must be addressed in repository service delivery.

### *Questions Considered*

In structuring their work, the task force identified a set of basic questions. These are:

- Looking across library engagement in the current repository environment, what are common concerns and questions?
- What are some of the strategies and options for service deployment?
- What can be said about where today's environment is headed?
- How should research libraries position themselves to succeed in digital repository service development?

- What issues need attention in the near term and which should ARL and its member institutions focus on?

### *Key Issues*

At this point, repositories are developing rather than developed. Repeated corrections, adaptations, and changes in direction characterize the experience of developers. To capture and convey a sense of the situation, the task force identified the following key issues.

**Building services around new content and old content in new forms.** Just a few years ago, many libraries were acting on a vision of repositories that focused on preprints and postprints of faculty publications and theses and dissertations. At the same time, digitization programs were producing modest numbers of files that required management. Early work with these forms of content revealed a further wealth of content requiring stewardship. We now understand better that institutions produce large and ever-growing quantities of data, images, multimedia works, learning objects, and digital records while mass digitization has launched a new scale of digital content collecting.

**Engaging with a larger networked environment.** The 21st century networked environment militates against managing repositories as isolated collections. Part and parcel of the multi-repository environment is the need to design repositories in ways that allow them to participate in higher-level, cross-repository services. For instance, search and discovery services are dominated by approaches that function across many repositories. As content ranging from raw data, analyses, simulations, research reports, meta-analyses, etc. are generated and likely stored in various repositories, repository services must be developed in ways that allow complex relationships between content in different repositories to be maintained and, likely, for much content itself to move back and forth.

**Attending to the "Demand Side."** Digital repositories are as much about users as they are about content, so the development of high-value repository services requires understanding user needs and capabilities. Repositories are intended to serve many kinds of users, but it is not necessarily obvious what this requires. Acknowledging and engaging with demand-side issues shifts the focus from building repositories to delivering services, and service development is a dynamic, heuristic process of identifying the diverse and evolving sets of demands and constraints posed by heterogeneous user groups. It is often initiated and always profoundly shaped by the demand side of the service equation.

**Sustainability.** Balancing needs, benefits, and resources is essential for service success. Sustainability is not merely about money; it is about organizational commitment and the ability to build persistent collaborations to address the ongoing needs for repository services and infrastructure. It is also about the beneficiaries of investments in repositories and understanding the benefits they receive. It is not clear that services are sufficiently developed to effectively project future needs let alone the resource demands they will pose. Thus, sustainable service programs need the ability to understand where they are succeeding and the flexibility to adapt as new opportunities emerge.

### *Common Strategies*

Libraries can draw on a variety of strategies for service deployment. This report explores three general strategies that are commonly used successfully to support repository services: in-house development and deployment, collaborative or partnering approaches, and contracting for services. Many libraries are using a mix of these strategies and may transition functions and services between them as time passes.

Numerous factors affect the choice of deployment strategy but, ultimately, it is related to the type of content to be housed in the repository, local resource



configurations, and the kinds of services needed. The latter are strongly conditioned by a host of institutional factors and will vary between institutions and over time. Fortunately, the repository environment has reached a point where it is increasingly practicable to do effective needs assessment and planning.

Existing repositories focus on varied types of content, offer diverse service regimes, address different users, and leverage a range of resource models. Cornell's arXiv program demonstrates an in-house development approach, HathiTrust and Ontario's Scholars' Portal programs exemplify different collaboration strategies, and the California Digital Library's publishing services incorporate outsourced repository services from BePress.

### *Horizon Analysis*

Surfacing common issues and considering common strategies provides some basis for recommending focuses for action, but an analysis of relevant environmental trends adds another important element. To investigate relevant trends and surface strategic issues, the task force developed a Horizon Analysis. The analysis considers four arenas within the digital repository environment: library users, the general technology environment, library collections and services, and key policy developments at the national and institutional levels. Taking the reference point of the year 2015, the Horizon Analysis paints a picture of near-term directions for change, and points to opportunities for ARL and member libraries to promote particular standards or national policy directions and engage in meaningful planning.

### *Roles for Research Libraries*

Looking at what has been learned by early repository service developers and considering the trajectories of external trends, it is evident that despite the varied funding and resource challenges faced by research institutions, delivering repository services is a crucial function of research libraries. The following issues are arenas where ARL and its members should focus attention and effort

to enhance research library roles in delivering digital repository services. This will often require collaboration within the community but also with other stakeholders active in the repository arena. Research libraries may not be able to lead in all of these areas, but should at least be seeking to make contributions to addressing current and emerging opportunities and concerns in each issue area.

- Develop a deep understanding of content users' and creators' needs to underpin the development of repository-related services.
- Apply a life-cycle management framework to guide development and evaluation of services and policies.
- Articulate a compelling value proposition for repository-related services to justify investing resources, promote partnerships, and address sustainability concerns.
- Integrate into emerging services the diverse content collections that have accumulated and will continue to arise outside of library-managed repositories.
- Participate actively in shaping the technology of repositories, particularly the mechanisms by which repositories make services possible.
- Negotiate the significant uncertainties existing in the current rights environment and build a broader consensus about the appropriate rights environment needed to support the research enterprise in a digital environment.

### *Call to Action*

Repositories are rapidly becoming ubiquitous in research institutions and libraries need to play an active role in service development. Even where libraries are reassessing their service portfolios in response to budget reductions, each needs to be developing expertise in this arena to participate in shaping these essential services. Diverse experiences with seed collections will deepen understanding of user needs, encourage experimentation

with different organizational frameworks, test different business models, and clarify options for managing technical infrastructure, making it possible to promulgate best practices and tackle system-wide issues. Important actions that research libraries should undertake include the following:

- Build a range of new kinds of partnerships and alliances, both within institutions and between institutions.
- Base service-development strategies on substantive assessment of local needs rather than blindly replicating work done at another institution.
- Engage with key local policy issues and stakeholders to encourage institutional engagement with national and international policy issues.
- Develop outreach and marketing strategies that assist “early adopters” of repositories to connect with the developing repository-related service system.
- Define a scope of responsibility to guide the development of repository services for varied forms of content.

The Task Force members believe that because of their powerful potential to enable key work and enhance the effectiveness of functions across the research enterprise, research institutions cannot afford to do without repository services, even in difficult economic times. Researchers and scholars with access to a spectrum of repository services will possess a substantial advantage in conducting cutting edge research, delivering high quality teaching, and contributing valuable services to society.

Repository management will not be the sole purview of libraries, but libraries have key strengths and missions requiring them to undertake various roles in repository service development. This report is intended to present a useful perspective on the digital repository environment and inspire ARL member libraries and others to assess their views and plans for service development. It is also important for ARL to incorporate the report's findings into its program planning and engage in the issue arenas the task force has identified.

## Introduction and Framing of the Arena

---

Digital repositories are developing rapidly as a key element of research cyberinfrastructure. Even when research institutions are grappling with difficult budget decisions in the current economic environment, they need to have a strategy for providing repository services.

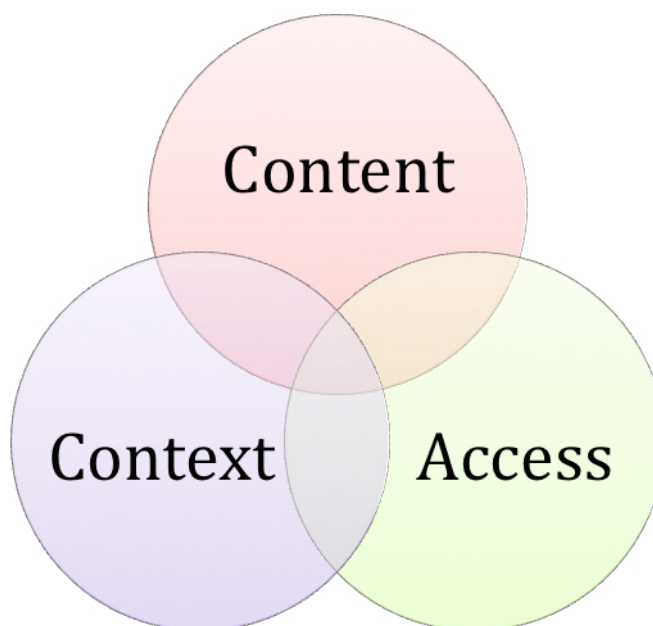
Libraries are making diverse contributions to the development of many types of digital repositories, particularly those housing locally created digital content — including new digital objects or digitized versions of locally held works. In some instances libraries are entirely managing a repository and its related services, but often they are working closely with other stakeholders at their institutions to jointly develop repository services.

The Association of Research Libraries (ARL) Digital Repository Issues Task Force was charged “to evaluate trends, contextualize repository activities among ARL libraries, and recommend leadership roles and activities for ARL.” It embarked on its work at a time when many formative repositories have been created, making it easier to begin to identify dynamics that will shape repository development and highlighting key issues and opportunities facing repository developers. With experience, pioneers are developing a clearer sense of vision, a deeper understanding of what might be achieved, and an awareness of where focused effort could make a difference.

In considering the repository landscape, technological considerations may leap first to mind as issues of great salience, but the task force members agreed they (and ARL) were best situated to focus especially on repository services — not machine services, but ser-

vices to authors, contributors, and users, particularly of university-created content.<sup>1</sup> For instance, some types of repository services provided by research libraries include long term archiving and migration of content, dissemination and access management, metadata and format management, search and discovery tools, publishing, data mining, etc. Repository services can also be deployed across a nearly unlimited diversity of content ranging from preprints of articles to born digital primary source materials, research data to university records, or software to video.

The experience of the California Digital Library (CDL) with its E-Scholarship repository shows the power of a service perspective. E-Scholarship was initiated as an open-access, postprint-oriented institutional repository program. Although considered unusually successful due to the relatively large numbers of digital works it has accumulated, a faculty survey<sup>2</sup> revealed that few of them were aware of the repository and many of those who did know about it lacked a clear understanding of what the repository offered. After a period of analysis and planning, CDL decided to refocus its program based on a redefinition of success emphasizing service delivery and service quality. The redesigned program will emphasize two related services: E-Scholarship and UCPubS that are designed as a service suite and are oriented to faculty and campus-based publishing. Based on ongoing user study, the interfaces for repository infrastructure are being redesigned to more effectively communicate and deliver publishing and dissemination services within a branded environment. Despite general budget reduc-



**Figure. Elements of Digital Repository Services**

Figure courtesy of Lars Meyer

tions, a position has been created to focus on marketing services and aligning them directly to campus and faculty needs. And, a new relationship with the University of California Press underpins an integrated set of services supported by both library and press resources, offering users a range of options that can be matched to a particular situation. For instance, the Press can deliver editing, printing, and distribution for monographic types of publications<sup>3</sup> while the library can provide a range of journal publishing and postprint dissemination services.<sup>4</sup>

The shifts in the CDL's focus from collection building to service delivery illustrate the ongoing evolution of the library's role in dissemination and publishing. As research institutions take on a greater responsibility for disseminating research and move into providing publishing services, typically through their libraries, these services are commonly intertwined with repository development and service delivery.<sup>5</sup>

Libraries have pioneered in developing institutional repositories, but these only represent a subset of the broad range of repository services they are providing. Institutional repository services expose many issues typical of those presented by services supporting a broader

variety of content types, contributors, and users, but libraries can draw on a broader range of experiences to shape digital repository service development. Research institutions need repository services that address a much broader range of research products, including research data.<sup>6</sup> Repository service development also draws on traditional research library strengths in working with institutional records and other kinds of special collections.<sup>7</sup> Therefore, the task force employed a broad conception of digital repository services rather than defining a narrower set, such as institutional repository services.

Repository services are built upon a foundation of content, context, and access. They rely on thoughtful development of each.

The early institutional repository movement emphasized access and in many instances institutional repository service development has struggled to build the content element. On the other hand, repositories coming out of digitization programs tended to initially emphasize content and continue to struggle with context issues such as metadata creation and standards support. Jointly, context and access issues underlie many of the emerging challenges of supporting discovery services.

## Questions and Frames

In structuring their work, the task force identified a set of basic questions. These are:

- Looking across library engagement in the current repository environment, what are common concerns and questions?
- What are some of the strategies and options for service deployment?
- What can be said about where today's environment is headed?
- How should research libraries position themselves to succeed in digital repository service development?
- What issues need attention in the near term and which should ARL and its member institutions focus on?

In examining these questions, the task force used four frames to organize their work: the user environment, the technology environment, library collections and services, and the policy environment.

The user environment encompasses a range of "demand-side" issues — user needs, user skills, and user expectations. Users are not just the people who interact with repository content after ingest, but the content creators or institutions that contribute to content creation. The expectations, skills, and capabilities of all of these users will and should substantially shape the development of repository services.

The task force members believe a number of venues exist for tackling technology questions, so where the task force considered technology issues, it was largely through their intersection with the other three frames. While not emphasizing technology issues, the general technology environment will inevitably influence the design of repository services. Therefore, the task force did consider relevant technology trends.

Repositories demand substantial reconfiguration of our thinking about collections and services in the library context. Digital is not merely a new mode for collecting

and disseminating the kinds of collections traditionally managed by research libraries. The digital age has unleashed a torrent of new kinds of content generated by the wide range of activities in which research institutions engage. In addition, it requires libraries to play a leading role in converting much of the content in their collections into digital formats.

A great many questions about technology deployment and service development come down to issues of policy. These are expressed at a number of levels, from the national level where laws express policy decisions to a variety of lower levels, including the local. The laws governing ownership, preservation, and use of works constrain and shape a great deal of what can be done with repository services and who can work with various kinds of content. Service development is also shaped by a variety of contracts and compacts between institutions and their faculty, as well as with external content providers. Large amounts of public money underwrite the work of research institutions and receipt of those funds carries growing expectations regarding the management of the products of research particularly. Resource constraints inevitably demand that a range of policy decisions be made regarding what content can be housed in repositories, what kinds of curation can be performed and what kinds of services can be developed. As repositories develop in collaborative and cooperative environments that place a premium on interoperability, addressing and aligning responses to a wide range of policy issues — from legal to contractual to institutional — looms large.

The task force used multiple approaches to gathering information and developing its recommendations. In identifying common concerns and questions facing repository services developers, task force members contributed their own experiences working with diverse repositories and supplemented these with targeted interviews of managers of additional exemplars of repository development. Consequently, much of the report is contextualized with examples drawn from these repositories. Similarly, some common strategies for developing reposi-

tory services are identified, characterized, and illustrated in this report.

A key exercise shaping the task force's discussions was a "Horizon Analysis" that collected its members' perceptions about the likely environment for service deployment in 2015. By looking out a short distance in time, it was easier to identify key trends and critical developments that, for better or for worse, are likely to both drive and constrain digital repository development. As a representation of the collective perceptions of a group with substantial experience with a first generation of digital repositories, the Horizon Analysis was enormously helpful in identifying those problems most needing attention.

Further sources for this report include information provided by other organizations on their activities relating to digital repository development (summaries appear in the appendix) and a variety of published sources. The bibliography presents a selection of publications the task force members found particularly informative.

The rather high level discussion of the key opportunities and issues presented in the report reflects a recognition that digital repositories are currently in a very early and dynamic state of development — one that encompasses diverse stakeholders, content types, and emerging service capabilities. Yet, it is vitally important for all research libraries to be engaging with digital repository development projects in some fashion. The report conveys a wide variety of options for libraries working in the arena, reflecting their belief that experimentation and first hand experience will be crucial to libraries at research institutions. Digital repositories are a fact of the twenty-first century research environment and libraries are developing experience and expertise both

with content management and service development. Significant engagement in repository service development will be essential if libraries wish to remain viable and vital in their contributions to the enterprise of research and scholarship.

## Notes

<sup>1</sup> An early and still valid articulation of the core service function of repositories is that provided by Clifford Lynch in, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, no. 226 (February 2003): <http://www.arl.org/bm~doc/br226ir.pdf>.

<sup>2</sup> University of California Office of Scholarly Communication and California Digital Library eScholarship Program. 2007. "Faculty Attitudes and Behaviors Regarding Scholarly Communication: Survey Findings from the University of California." Berkeley, CA: University of California. <http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf>.

<sup>3</sup> The University of California Press is reserving its imprint apart from the new service suite.

<sup>4</sup> This account is based on interview of Catherine Mitchell, Director, eScholarship Publishing Group.

<sup>5</sup> See Karla L. Hahn. "Research Library Publishing Services: New Options for University Publishing." Washington, DC: Association of Research Libraries, March 2008. <http://www.arl.org/bm~doc/research-library-publishing-services.pdf>.

<sup>6</sup> See ARL E-Science Task Force. "Agenda for Developing E-Science in Research Libraries: ARL Joint Task Force on Library Support for E-Science Final Report & Recommendations." Washington, DC: Association of Research Libraries, November 2007. [http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf).

<sup>7</sup> ARL's Special Collections Task Force will be releasing a report in early 2009 addressing new roles and priorities for research libraries.

## Common Issues and Choices

---

The repository environment is a key element of the cyberinfrastructure developed in the past decade and, like other elements, repositories are developing rather than developed. Repeated corrections, adaptations, and changes in direction characterize the experience of developers to date. To capture and convey a sense of the situation, the task force identified the following key issues and developed cases of existing digital repositories that exemplify currently employed strategies.

### *New content plus old content in new forms*

One of the conspicuous features of the emerging repository environment is the diversity of content types involved. Just a few years ago, many libraries were acting on a vision of repositories focused on preprints and postprints of faculty publications and theses and dissertations. At the same time, digitization programs in libraries were producing collections of modest numbers of files. These early efforts to create repository services revealed a further wealth of content that potentially requires stewardship. Research libraries are more broadly conceptualizing repository services now as they understand better that institutions produce large and ever-growing quantities of data,<sup>8</sup> images, multimedia works, learning objects, and digital records while mass digitization has launched a new scale of digital content collecting. Libraries often have established, if limited, roles in working with content such as course reserves and university publications. Twenty-first century institutions now require new kinds of services to manage all sorts of unique content that have enduring value.

Beyond the obvious issues relating to technology that arise around archiving diverse kinds of content, additional issues arise out of the various ways content is created, collected, or converted into digital form. Content can be text or images that are born digital, or digital representations of works originally produced in traditional media — books, articles, photographs, or sound recordings, to give a few examples. Digital content with print surrogates (and vice versa) raise a host of issues around coordinating management of both content forms, as well as versioning.

Content forms that rarely required institutional management in the past can now benefit from organized archiving and curation services. Instrumentation of various sorts is generating large amounts of digital data useful for various research projects, but largely unprocessed. Selecting new kinds of content for storage, and accounting for the stage in the research process during which it is generated, become important.

The motivations for creating the content also affect key decisions about management and service development. Some content may be intended for long-term use from the outset — traditional publications, for instance — while others may be valuable for a comparatively finite period of time. Or it may be difficult to tell, initially. Two examples of institutionally generated and locally valuable forms of digital content, learning materials and institutional records, provide a sense of the range of issues that shape service development.

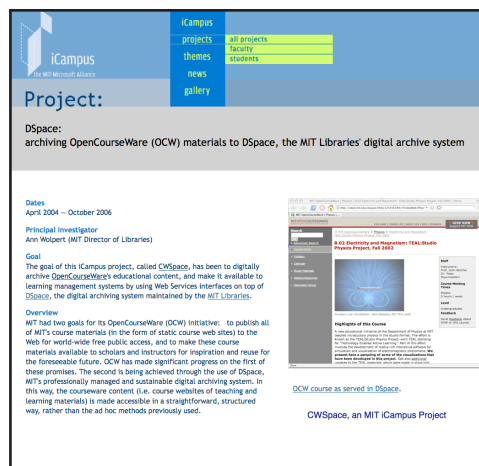
Course-related content is accumulating rapidly on university systems. Course materials are the products of

major organizational investments and often have ongoing value, yet models for providing repository-based services are few and require research libraries to build or strengthen relationships with other campus stakeholders charged to provide instructional support.

MIT's OpenCourseWare (OCW) program takes the innovative approach of organizing the dissemination of course materials using a publishing model. Working with this model, MIT Libraries' have developed services that meet the OpenCourseWare program's need to maintain a focus on production and maintenance of current content but ensure that earlier versions of superseded materials are archived through the CWSpace project. The Libraries leveraged their initial relationship with the OCW program, based on provision of metadata services, to build an understanding of the program's archiving needs and then formulated appropriate services to meet the needs of the program and a variety of users who continue to find value in the superseded materials.

Learning materials are just one of a larger group of "institutional assets," items that have financial, historical, or intellectual value to the institution, and thus need to be documented and preserved. Such assets require many stewardship decisions but need not be treated homogeneously; in addition to retention times, levels of metadata, user access, expected lifespan, add-on services offered (e.g., extraction, analysis, recombination, data mining) might all vary. Many repository services distinguish in their strategies and policies how they provide services to the academic functions of research institutions and the administrative functions. Libraries at research institutions play particularly varied roles in supporting administrative functions. Consider the case of administrative and archival records related to the institution itself: files and records from financial, personnel, legal, admissions, planning, research administration, public relations, faculty processes, and all the other areas of university operations. Many research libraries house traditional university archives and may be positioned to integrate digital records into existing

## The Example of MIT's CWSpace



MIT OpenCourseWare (OCW) is a Web-based publication of MIT course content. It is open and available to the world and is a permanent MIT activity. Utilizing a publishing model, OCW regularly updates, replaces, and deactivates content. MIT Libraries provide OCW with archiving services to preserve de-selected content through its CWSpace program.

MIT has two goals for its OpenCourseWare initiative: to publish all of MIT's course materials (in the form of static course web sites) to the Web for world-wide free public access, and to make these course materials available to scholars and instructors for inspiration and reuse for the foreseeable future.

CWSpace archives superseded materials from the OCW publishing initiative, along with their metadata, providing ongoing access to the content. This content is the most heavily used collections of the Libraries' DSpace program and is in regular demand by MIT faculty for course development.

Based on information provided by MacKenzie Smith, Associate Director for Technology, MIT Libraries.



archives. Other institutions have established different organizational structures, but many are turning to libraries for assistance in developing digital repository services.

Institutional records may be critical for legal requirements, mandated public access, business continuity, or historical purposes. Which records to keep, how to keep them and where to draw the line may depend on which criteria or audiences are considered most important.

As digital records accumulate, they pose such questions for digital records management as:

- How should record types be defined and categorized? University records increasingly include dynamic forms such as those existing in relational database environments. These often contain multiple files and dynamically updated content, rather than static objects.
- Many universities lack systematic records management and disposition policies governing the retention of administrative records, leaving individual offices to make decisions about backup, storage, and deletions in a decentralized and undocumented manner.
- In universities that do have such policies, responsibilities may be decentralized and/or digital content may not be assigned to the university archives. There may be tricky political conflicts among university offices that feel they have the oversight for records management, including not only the library but IT, business services, general counsel, institutional research, even facilities.
- In the digital environment, backup of administrative computing (when it occurs!) may be handled in a periodic batch mode and may not incorporate formal curation such as the ability to restore records in original form, use of metadata for record tracking and identification, or monitoring the digital integrity of the objects.
- Complex privacy and confidentiality restrictions on university records may be in place,

yet also run up against public disclosure requirements. Institutions may be reluctant to develop centralized repositories because of the difficulty of monitoring access or the security vulnerabilities that might result from aggregating the information.

- Heavy use may be made of digital audio, video, and social networking sites for university publications and communications, yet these are often seen as ephemeral objects and may be retained, if at all, only in working files for short periods of time, unlike print counterparts of an earlier era that might be collected in the historical archives.
- Poor "version control" is common for digital records since objects are rewritten and reused without keeping separate distinct editions, and yet, for the documentary/archival audit trail, version and date are essential.

As daunting as it may be to consider the diversity of such issues, research libraries have substantial expertise in content management and opportunities to support high-value institutional content and to complement repository services being developed to serve campus research. Particularly where these activities draw on existing relationships, capacities, and expertise, libraries are likely to meet with success in assuming a broader role in supporting content such as course-related assets or institutional records.

### *Engaging with a larger networked digital environment*

Although many repository services focus on locally created content, the twenty-first century networked environment militates against managing repositories as isolated collections. For instance, search and discovery services are dominated by approaches that function across many repositories. Thus for many repositories, offering capabilities to participate in shared discovery services will

be more important than developing sophisticated local search capabilities.

Part and parcel of the multi-repository environment is the need to design repositories in ways that allow them to participate in higher level, cross-repository services. Support for shared service layers and API's is also sorely needed, as the recently intensified conversations between the DSpace and Fedora initiatives show. Similarly, support for standards like the OAI's protocol for metadata harvesting and object reuse and exchange specification is integral to successful service development. Some Web 2.0 styles of functionality, such as "mash-ups" also rely on cross-repository capabilities.<sup>9</sup>

Many repositories are at only the beginning of their growth curve and, while some are poised to achieve substantial scale, it seems likely that repository proliferation will be the norm for quite a while. Efforts to manage diverse content in a dynamic and immature service environment may favor repository services tailored to the characteristics of well-defined collections of content and users with relatively clear needs. As long as repositories are exposing high quality metadata, maintaining interoperability, and supporting data exchange, this is a good strategy, at least in the short term.

Some of the need to access, share, and link content across repositories will be driven by the broader array of digital content generated throughout (or across) the research cycle. Content ranging from raw data, analyses, simulations, research reports, meta-analyses, etc. will often be stored in various repositories. In this emerging environment it becomes important to develop services able to ensure the integrity and authority of content while also maintaining and creating appropriate relationships between various items and their copies or different versions.

### *Attending to the "Demand Side"*

Digital repositories are as much about users as they are about content, so the development of high value repository services requires understanding user needs

and capabilities. Repository creation is often motivated by the existence of content that requires some form of management and curation. From the first stages of service development, repositories are intended to serve users, but it is not necessarily obvious what this requires. Acknowledging and engaging with demand-side issues shifts the focus from building repositories to delivering services.

Typically, developing new services is a dynamic, heuristic process of identifying the diverse and evolving demands and constraints posed by heterogeneous user groups. While to a casual observer it may look like a service is launched and users appear, in fact, service development is often initiated and always profoundly shaped by the demand side of the service equation. Operating within a complex and changing environment, repository services have to meet many different constituencies' varied needs. Balancing competing demands of different user groups typically requires policy development as much as technology development.

Libraries often interpret "user" in a service context as conveying the human searcher or reader — a person discovering content and applying it to some individual need. This is far too simple a picture, however, to use in planning and evaluating repository services. Demand also originates at community levels. Research institutions, disciplinary or research communities, and libraries themselves may all function as demand drivers. Each community will have well-established practices, tools, expertise, and expectations. Other service users include mediators of various kinds who work to recruit or collect content — perhaps as part of a formal editorial role or perhaps in an effort to advance some community's research activities. A different sort of service consumer in the networked environment is other IT systems that draw on repository content for machine processing or computational analysis underpinning a different set of service functions, for instance text mining, or discovery.

Demand drivers also arise from the needs of content creators. These may be the same individuals who

are also discovering and using repository content, but the content creator role presents different implications for service demand. For instance, content creators may have particular concerns about access management, rights allocation, version handling, metadata creation, and long-term stewardship. Collectively, potential users of repository content share concerns about discovery and retrieval, permitted uses, and persistence of content access. Any human users tend to be interested generally in ease of use and increasing their ability to integrate repository services with their existing tools and workflows. Thus, effective needs assessment requires developing an understanding of user workflows. These differ substantially between disciplines, but, even within more focused research communities, shared sets of practices can arise. General ease of use, of course, has a large impact on the success of repository services. But they have to be designed for close integration with varied non-library work activities occurring on the desktop. An additional trend with strong implications for service design is the growth in collaborative research and scholarship with the result that repository services need to support group as well as individual work.

Demand is also driven by the various reward systems that motivate or inhibit service development. Academic researchers are typically highly responsive to the tenure and promotion regimes to which they are subject. All individuals and many institutions are motivated by prestige-based reward systems and the inherent competition they engender. Funding systems can also shape institutional and individual needs. Of course, these can overlap and interact, as well.

In the digital networked environment, repository services can blur the library/publisher distinction while extending both roles. Content curation and stewardship roles have traditionally fallen along different lines than content creation and publication, with libraries responsible for stewardship and access and "publishers" organizing the creation and distribution of content, increasingly within a market context; repository services can cover

either or both roles. This is one reason why the range of potential users for repository services is so broad. A repository may initially focus on one of these two functions and over time expand into the other. Alternatively, repository services can be constrained by policy decisions that promote or discourage the development of an emphasis on only one set of roles. For instance, repositories may focus on previously published content and focus exclusively on assuming curation and stewardship functions.

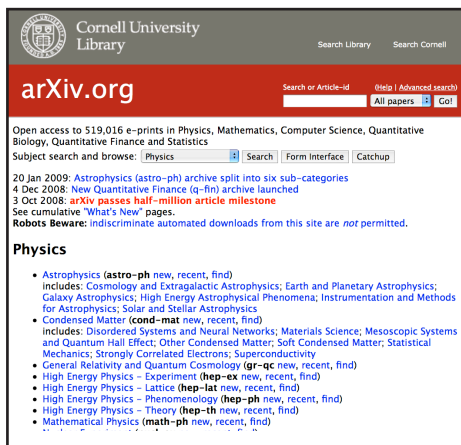
Wherever repository services fall on the dissemination/publication spectrum, repository service developers are confronted with issues related to users' needs for ensuring the authority and integrity of works as well as the ability to distinguish between different versions of works. Both content creators and content consumers need to be able to assess the place of any particular version in the broader authoring process, be assured of the integrity of the content submitted or retrieved, and be able to readily discover related works.

In the face of the diversity and complexity of demand-side issues, the adoption of a strategy of market development is perhaps key to integrating them into service development. Successful service development is more likely when an initial launch is based on thoughtful exploration of user issues (broadly defined). This should be followed by regular user feedback with an expectation that service design needs to be flexible and adjust as experience is gained. Successful service programs must understand where they are succeeding and possess the flexibility to adapt as new opportunities emerge. Collecting information on an ongoing basis about who is using services, what is being used and how it is being used positions service programs to attract the resources they need on an ongoing basis.

### *Sustainability*

As repository services become a crucial part of the cyberinfrastructure underpinning research and scholarship, new strategies for financing and organizing them are needed. Digital repositories today are immature;

## The Example of arXiv



Over more than a decade, arXiv has transformed the research communication infrastructure of multiple fields of physics and plays an increasingly prominent role in a unified set of global resources for physics, mathematics, and computer science. Paul Ginsparg originally developed the digital repository in 1991 as an archive for preprints in physics. The repository moved to Cornell with him in 2001, and is now a collaboration between the Cornell University Library and Cornell's Information Science Program. The Library is responsible for arXiv's maintenance, while research and development of the repository is handled by Information Science.

arXiv is an international initiative and involves collaborations with US and foreign professional societies as well as other international organizations. Most scientists and researchers who post content on arXiv also submit it for publication in traditional peer-reviewed journals. Famously reclusive Russian mathematician Grigori Perelman's decision to post his proof of the 100-year-old Poincaré Conjecture solely in arXiv underscores the repository's increasing importance and its role in transforming scholarly communication.

As of May 2008, it contains over 490,000 articles; over 60,000 new submissions and 55 million full-text downloads are projected for calendar year 2008. With roughly 100,000 distinct users per day worldwide, it is the most heavily used online service of the Cornell University Library. Currently, not including R&D efforts that mainly take place within Information Science, the service staff includes 3.5 FTE staff with a budget of approximately \$300,000 provided by the Cornell University Library.

The essential success principle for arXiv is its being firmly embedded in the research workflows of these subject domains. Through Paul Ginsparg's leadership with his roots in both academic and IT communities, the service consistently focuses on the disciplinary cultures represented in the digital repository and community needs. Yet, while the underlying technology for arXiv has been updated throughout its 16-year history, arXiv requires significant internal re-engineering to support an evolving technological landscape, increased growth and use, and to ensure the sustainability of the service. For example, access to the archived content, including supporting data and other attachments as well as text, will facilitate re-use and re-purposing of underlying data and information in order to further advance knowledge.

Based on information provided by Oya Rieger, Associate University Librarian for Information Technologies, Cornell University.

they draw upon emerging technologies, manage novel forms of content, and are just beginning to explore the service demands arising from the new capabilities of the digital age. Sustainability is a pervasive question among service developers and one where experimentation and a willingness to learn from both success and failure are particularly helpful. In the current environment, it is not easy to project what future resources may be required to support repository services, or what sources of support could be used effectively over the long term. The emerging global economic crisis makes sustainability questions particularly substantive.

Repositories are growing rapidly, but not as quickly as the bodies of content requiring archiving and management, making capacity development a significant issue. Typically, a repository may be launched to meet a particular service need, but, as the utility of repository services becomes clearer, demand increases as services are refocused to better match user needs. Or, the original focus of the repository may mushroom as the volume of digital content being used begins to grow. Service developers must balance investment in development and marketing with the demands of service growth.

Perhaps the most fundamental issue with regard to sustainability questions is just how little is known about repository services and content management. While various experiments have been set in motion, libraries are only starting to build a collective understanding of what repository managers and administrators are learning. It is not clear that services are sufficiently developed to effectively project future needs, let alone the resource demands they will pose.

Digital repositories require ongoing content curation to ensure that current content remains usable and valuable into the future. Much digital content is unique content, in the sense that only a single institution may be able or willing to take responsibility for its management. This uniqueness may be incompatible with traditional strategies for printed publications that rely on redundancy or some kinds of shared investment to provide stabil-

ity and ensure continuity of curation functions. Curation costs may vary across the life cycle of content making it even trickier to predict ongoing resource demands.

Digital preservation stands out as an area that is starting to receive the kind of economic analysis needed to allow productive conversations about long-term support for a wider array of repository services (see the recently released report by the Blue Ribbon Task Force on Sustainable Digital Preservation and Access<sup>10</sup>).

As research libraries embark on repository service development, they enter a brand new business in many ways. Existing service models, organizational structures, and staff expertise are often not directly transferable to the task of repository-based service deployment. While repository services present libraries with attractive opportunities to develop roles that will become increasingly mission-critical for the research enterprise, service development requires libraries to engage in a larger environment, one requiring new partnerships with various stakeholders within an institution and, in many cases, between organizations.

Herein lie issues of developing value propositions for repository services. Very little formal measurement or even articulation of the value proposition for repository-related services exists yet. The relative importance of providing ongoing support for repository services will inevitably be conditioned by the real and perceived benefits that result from repository-based services. Demand for stewardship of digital content is strong, but the actual benefits of repository functions are largely only notionally understood. In addition to direct value added by services, more proximal benefits like gaining competitive advantage or avoiding opportunity costs affect the value proposition for services for a given institution. Both in seeking new opportunities and making explicit choices about deploying resources, libraries need to consider both service inputs and benefits.

Many institutions have some first-hand experience to apply, but beyond resource inputs and the development of organizational relationships, relatively little has been

described in the way of outcomes from repository service use. As research libraries face difficult choices driven by real reductions in their resource bases, they must assess the centrality of repository services to the research and teaching functions of their institutions, convey their importance to decision-makers in the broader organization, and let go of lower priority activities, if necessary. At bottom, sustainability is not merely about money; it is about organizational commitment and the ability to build persistent collaborations to address ongoing needs for repository services and infrastructure.

## Notes

<sup>8</sup> The ARL E-Science Task Force report, "Agenda for Developing E-Science in Research Libraries, ARL Joint Task Force on Library Support for E-Science Final Report & Recommendations," Novem-

ber 2007, offers extensive analysis of data management issues in the context of science research. [http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf).

<sup>9</sup> The term "Web 2.0" is difficult to apply effectively as it is diversely used to refer to technologies that facilitate new forms of interaction between individuals (a la Facebook) that are less relevant to repository services, to architectures that aggregate machine services through the web to provide what appears to the user to be a single source, or to services that facilitate user "mashups" by allowing users to select and aggregate data from different sources. Rather than referring to Web 2.0, throughout the report, the task force prefers to simply describe the relevant functionality directly.

<sup>10</sup> Brian Lavoie, Lorraine Eakin, Amy Friedlander, Francine Berman, Paul Courant, Clifford Lynch, and Daniel Rubinfeld. 2008. "Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation. Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access." [http://brtf.sdsc.edu/biblio/BRTF\\_Interim\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf).

## Strategies for Service Deployment

---

Engaging in digital repository development is a high priority for research libraries, one that will be essential for maintaining their value in research organizations. Although there are many routes to engagement and many strategies for service deployment, this report explores three general approaches to supporting repository services: in-house development and deployment, collaborative or partnering approaches, and contracting for services. Many libraries are employing a mix of these strategies and may transition functions and services between them as time passes.

Choosing among strategies requires consideration of a wide range of factors. Ultimately, the type of strategy deployed depends on the type of content to be housed in the repository, local resource configurations, and the kinds of services needed. The latter are strongly conditioned by a host of institutional factors that will vary between institutions and over time. Fortunately, the repository environment has reached a point where it is increasingly practicable to do substantive needs assessment and planning. Existing repositories focus on varied content types, offer diverse service regimes, address different users, and leverage a range of resource models.

Even with many models from which to learn, planning for flexibility is imperative, since demand is often delayed into the future and content is likely housed (or at least produced) in a dispersed mode. An appropriate strategy may be to plan small initially but expect to respond to growing demand. Further, it can help to think of repository service development as a business development activity. Instead of just launching reposi-

tory services as add-ons to existing activities, a business development approach involves environmental scanning, investigation of models, thoughtful consideration of alternatives, and assessment of available and potential resources. Business development is iterative and services evolve throughout the deployment process.

Three key factors shape deployment or development decisions: the desired service regime, policy, and resources. Identifying the desired service regime depends on the potential sources of content and demand for user services. These considerations should be framed by the institution's mission, goals, and top priorities. Repository service development needs to align with the resources and investments that are already in place.

The local policy regime should also be considered early in the process. Thought should be given to institutional rights management practices and policies, privacy and security requirements, reporting systems, and strategic commitments. If repository development requires shifts in the local policy regime, these need to be identified and the work begun as early as possible. This reduces the risk of becoming embroiled in conflicts between services and policies.

Options for infrastructure development tend to be constrained by local and historical factors. Choosing among the options will depend on available sources of funding and expertise; the library's span of control; historical relationships; and the timeframes for service development and requirements for sustainability that confront the library.

### *In-house development and deployment*

This is the paradigmatic strategy of institutional repository development, but it can make sense for other types of digital repositories, too. Open source applications are available (e.g., DSpace or Fedora) making it possible to get a repository up and running with quite modest resource investments. ARL member libraries surveyed in 2006 reported that nearly 80% of those with institutional repositories (a subset of digital repositories) were using a local deployment strategy and more than half of them were using DSpace.<sup>11</sup> Open source software lets libraries begin small-scale repositories to explore service development. Alternatively, some content “needing a home” can be housed quickly by deploying existing applications and local servers.

Other factors may favor in-house deployment. When there are substantial issues around rights management, privacy, or security, it can be important to retain content and ensure access control locally. In-house development may be the best option for work with cutting edge content or digital objects. Highly individualized service needs may make internal development the fastest and most effective approach.

Where local deployment is coupled with local software development, an institution has maximum control over the repository design and management, but this is a highly resource-intensive strategy. Harvard's experience suggests that even with substantial resources, in-house development of a repository service faces a great deal of hard work with on-campus partners, in developing policies and in planning and coordinating the development of diverse service functions. One of the benefits Harvard has realized from their strategy has been the development of very strong relationships with some key users of repository services, particularly a wide array of content providers like campus museums and archives.

Staying in-house also doesn't necessarily ensure that digital repositories will be easily grafted onto existing library IT functions or traditional service delivery; in

### **The Example of Harvard's Digital Repository Services**

Part of a decade old Digital Library Initiative, the Digital Repository Service (DRS) provides a set of professionally managed services to store, preserve, and provide access to digital objects over time. Services are available to libraries and museums across Harvard, and are partially cost-recovered. The DRS has been developed in-house as part of a suite of independent but cooperating services operated by Harvard University Library to provide a comprehensive infrastructure for discovery, storage, and delivery of digital content. The DRS today contains over 70 terabytes of data, deposited by 46 different administrative units across the University.

Although staff have recently evaluated use of open source or commercially available repository platforms, a decision was made to continue to use and pursue further development of the locally-developed system. At this point, the cost of conversion and of integrating a new platform into Harvard's digital library infrastructure outweighs the advantages of using a shared platform.

The effort to develop and maintain the system has been significant. The repository system continues to evolve as the understanding of digital preservation increases and standards and best practices evolve. Further, there is an ever-growing range of materials to be managed by the repository, including records for print collections, botanical type specimens, audio, video, and art works, among others.

Reflecting the increasing importance of the repository to the library, Harvard has recently created a new position at a management level responsible for both the services of the repository and for digital preservation.

Based on information provided by Tracey Robinson, Head, Office for Information Systems and Wendy Gogel, Digital Project Program Librarian, Harvard University.



## The Example of Scholars Portal



Scholars Portal has its origins in the electronic journal services implemented by the University of Toronto Libraries in 1997. Subsequently these services were extended to seven Ontario universities on a fee-for-service basis. In 2002 the University of Toronto Libraries' service provider relationship was expanded to include all of the Ontario universities with start-up funding from the Ontario government to establish the infrastructure for what has come to be known as the Scholars Portal.

Services supported through Scholars Portal include curation, preservation, and access management for a range of content types including: electronic Journals (13.3 million locally archived articles from 8,300 journals from 20 scholarly publishers); index and abstract databases (more than 200 with over 150 million citations); research repository works (housed in a number of instances of DSpace); electronic book collections will soon include scholarly e-books (30,000 contemporary titles and 160,000 out of copyright titles) and research data (a numerical, statistical, and geospatial information service will soon support a wide range of information resources and, where possible, integrate these resources with other services).

Based on information provided by Carole Moore, Chief Librarian, University of Toronto.

many instances different skills, relationships, and expertise are required. MIT, a pioneering library in service development, has learned that the skills needed for managing classic ILS and related systems are somewhat different from those required for repository management. A library's existing expertise may not easily transfer to repository service support. Still, in-house development, if successful, can provide a great deal of experience quickly and ensures the development of local expertise.

### *Collaborations and partnerships — within institutions, across institutions*

Examples abound of collaborative and cooperative approaches to repository deployment. While they can function successfully for various content types and service models, collaborative approaches are far from a panacea; success requires good faith and investment from all the players. Where there is a shared commitment to the approach, there are substantial benefits. Early experiences suggest the strategy tends to work well where there are existing relationships to build on, particularly when considering extra-institutional collaborations. Several examples now exist of collaborative development of repository services within consortia.

Successfully addressing varied service regimes requires common agreement on needs and shared interest in the content. Preservation and access services for commercially acquired or locally created content, publishing services, and data curation are some examples of collaboratively developed repository services.

Two examples serve to illustrate cooperative service development, although collaborative elements occur in several other repositories described in the report. Scholars Portal, managed by the University of Toronto Libraries for the Ontario Council of University Libraries (OCUL), provides an example of an existing consortium deploying a common architecture and then fitting that architecture to diverse content types and service needs.

The Hathitrust Project that serves the Committee on Institutional Cooperation (CIC) and the University of

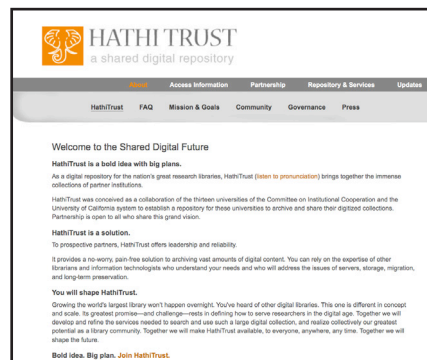
California, in contrast, is very focused on a single content type and operates with a well-defined and articulated purpose.

In both cases, all participating institutions contribute resources to the project, but service development is implemented by one institution, the University of Toronto in the case of Scholars Portal, or two, the University of Michigan and Indiana University in the case of HathiTrust.

HathiTrust and Scholars Portal have both confronted a key concern for multi-institutional repositories: balancing governance and funding. In the case of Scholars Portal, the service has evolved from a university-funded service using a model OCUL has adopted for sharing the costs of services. The program now has a Steering Committee, reflecting different constituencies within the consortium, that meets frequently and presents options and budget proposals to the OCUL Executive (some overlapping members) and full OCUL membership meetings, which occur in fall and spring. While this governance worked well initially, as the scope and focus of services have changed and the expectations for what are core services and what are opt-in services have evolved, cost sharing models have been a source of conflict. Further, as government funding has been provided for the implementation of new services, the OCUL members are facing new challenges in the ongoing sustainability of these initiatives.

In the case of HathiTrust's organization, one early and pivotal issue was finding a method to get sufficient feedback from partners without becoming bogged down by consensus-driven processes. Early documents described governance as a central issue and proposed exploring governance models through discussions with potential partners. HathiTrust's ultimate strategy was to create a single executive management group constituted by the library deans and CIOs of the lead institutions, along with an executive director, and have that group coordinate decision-making with a strategic advisory board.

## The Example of HathiTrust



HathiTrust is a large scale, jointly developed multi-institutional repository focused initially on digitized book and journal content coming out of large-scale digitization efforts, but plans to expand in the coming years to other formats and types of content. Released in October 2008, HathiTrust was initiated by the University of Michigan and Indiana University, and the founding partners are the remaining institutions of the Committee on Institutional Cooperation (CIC) and the University of California. It is seeking other research libraries as partners in the effort as well.

HathiTrust is designed as a very high capacity repository and launched with more than 2 million volumes available online, and adds hundreds of thousands of volumes per month. It has developed the infrastructure to preserve this content and, where permissible, provide access to it. (It is in the final stages of TRAC certification and is in conversations with CRL about a formal review.) To enhance discovery, it supports an API for online catalogs to annotate records with the availability (and level of access) of content, and regularly distributes metadata that documents content held in the repository.

Based on information provided by John Wilkin, Associate University Librarian for Library Information Technology and Technical and Access Services at the University of Michigan Library.

Financial issues faced by HathiTrust include finding effective ways for institutions to fund the initiative from permanent sources (rather than relying on grant funding) and convey to their internal constituency the value of contributing resources for the purpose of furthering a collective good (i.e., that this is a reasonable and not a competing way of meeting local needs). These sustainability concerns were addressed by developing the repository within a group of libraries and not as an external service. Despite obvious complexities raised by this strategy, this approach was chosen to ensure that research libraries, with their commitment to permanence of the human record, would be directly responsible for the effort.

Establishing shared agreement on goals and mission is an ongoing process for both Scholars Portal and HathiTrust. The Scholars Portal program continues to grapple with maintaining a clear vision as the Scholars Portal staff has grown, as the leadership within the OCUL and Scholars Portal management structures has changed, and as more people from the OCUL libraries have become involved.

Differing expectations must be regularly negotiated and reconciled. For instance, initially all of the Scholars Portal services were “core services” available to everyone. Recently, Scholars Portal has introduced some “opt-in services.” This divergence is driven by two factors: 1) interest of individual institutions to participate in new initiatives, and 2) the ability of individual institutions to pay for new initiatives. These two factors are often related.

There is also divergence in the definition of the scope and focus of the Scholars Portal services. Some feel that Scholars Portal should be focused on serving only the Ontario universities, yet the reality is that Scholars Portal is the de facto national provider of RefWorks services in Canada, and recent funding from the Ontario government has been provided with the expectation that some services will be extended to people throughout the province.

Intra-institutional collaborative service development is another common model, although one that raises somewhat different issues. There are many advantages to deploying collaborative strategies working with partners operating within the same institutional environment since repository development benefits from a rich understanding of the local environment — policies, needs, existing services, and resources. Multiple intra-institutional partners must still negotiate processes for decision-making and resource contributions, however.

It may be wise to avoid committing to an in-house service development strategy until the possibilities for intra-institutional collaborations have been explored. During the needs assessment phase of service planning, it is important to connect with various campus players who may control content, desire various repository services, have responsibility for key policy decisions, manage useful infrastructure, and so forth. Potential service consumers often can become key partners and contributors to service development.

Research libraries generally have close and long-standing collaborative relationships with academic administrative functions and faculty — reflecting the central priority typically given to faculty research at a research institution. They may be the most strategic partners to engage initially in the development of repository services. Alternatively, some institutions may have built close relationships with key units supporting teaching and learning and may find richer partnership opportunities and deeper institutional support for long term investment and key policy shifts by focusing first on instructional support functions. In addition, potential partners can include such units as: high performance computing, storage planning units, offices for research and grants management, deans of colleges and academic units, internal audit staff, university records, museums, or a university press, to name a few. Of course, not all of these partners must be considered or can work effectively in collaboration with the library, but it is important

to begin by being open to collaborations. Campus collaborations will naturally be most effective where there is overlap in the mission, goals, and top priorities of the potential partner and the library.

### *Balancing local and vended services*

Hosted services are available from many sources and may be a useful part of a repository development strategy. Repository development is too embedded in a local context and service environment for service provision to be fully outsourced, but fairly advanced services are available to support hosting. Two examples include the BePress Digital Commons product, which many institutions are using to provide institutional repository services, and OCLC's ContentDM, which is designed to support collections of a broader range of digital objects.

Hosted repository services can allow rapid start up and demand minimal technology support. However, contracting libraries will still need to make investments in service development — even vended digital repositories will not take care of themselves. The California Digital Library E-Scholarship Repository mentioned in the introduction, illustrates the tradeoffs. Initially staffed very lightly, the project relied on BePress to serve as their vendor and provide infrastructure development, while project staff focused on marketing and outreach. Over time, CDL staff concluded the interface offered by the vendor could not easily support the needed repository services and CDL is now developing a new interface in-house, in close consultation with campus users.<sup>12</sup> The CDL experience illustrates the need to assess services throughout their development phase and deploy new approaches as better understanding of “demand side” issues emerges. When establishing a contract for hosted services it is important to ensure that as the local environment evolves, the library has the flexibility to make changes in its vending strategies.

Several situations may reduce the advantages of outsourcing. When working with a vendor, the contracting library may make little investment in technical infra-

structure, but in turn has little control over development decisions that effect local service provision. As a deeper understanding of service needs develops, features or modifications of the hosting service may be required, but program managers may find they cannot obtain the changes needed. Outsourcing content hosting may be inappropriate for certain types of content, for instance when specialized management capabilities for privacy or security reasons are required. The need for scalability may also raise questions that should be considered at the outset. Vended service that seems affordable for managing modest amounts of content may become unmanageably expensive as demand grows and the content collection expands rapidly.

Hosting strategies will tend to work best when a library has a well-defined need and plans for the eventuality when content and metadata will need to be migrated out of the system — either to a local system or to another vendor, and has vetted in advance any rights management, security, and confidentiality issues. For smaller institutions with limited resources to do development work, outsourcing can be very efficient. In addition, larger institutions could offset their development costs by providing services to consortia or smaller institutions. The scarcity of commercial vendors for some kinds of repository services means there is a relatively open field for institutions to explore becoming service providers themselves.

### *The bigger picture*

The reality is that many research libraries can work with multiple repositories and a range of strategies. One library may use a hosted institutional repository service, participate in a consortial repository of licensed content, and manage a local repository of digitized images. Another may base its institutional repository services in-house and outsource hosting of its digital image collections.

Regardless of the deployment strategy, or combination of strategies undergirding repository services,

libraries have to develop their own local engagement. Even where initial investments are small, repository service development is the entry point for a new line of business for libraries, one requiring new and different kinds of organizational models. While services may form around existing relationships, repository services will raise new issues and require new investments. The technology issues inherent in repository services are different from traditional library IT expertise. Delivering repository services demands different skills, relationships, and infrastructure. To develop the staff needed, it will be necessary to tap into different networks of practitioners — experts in digital curation, for example — that will push the organization in new directions and reshape traditional functions.

It is important to be mindful that libraries are only one part of a diverse array of players that are launching

and managing repository services. For the library community to participate in a broader arena of repository development, their claims to a place at the table must be underpinned by experience delivering services within their local context. While repository services often extend well beyond the local institution, they benefit from, and cannot succeed without, the kind of deep interactions libraries can create with local constituencies.

#### Notes

<sup>11</sup> Bailey, Charles W. *Institutional Repositories*. SPEC Kit 292. Washington, DC: Association of Research Libraries, July 2006. <http://www.arl.org/bm~doc/spec292web.pdf>.

<sup>12</sup> Interview with Catherine Mitchell.



## A Horizon Analysis

---

Task force members engaged in a Horizon Analysis to surface strategic issues and clarify trends. They described key environmental conditions they believe are likely to prevail by the year 2015. While painting a picture of some near term directions for change, the task force looked for areas where developmental trajectories appear to be on-course and where there are opportunities for ARL and member libraries to promote particular standards or national policy directions and engage in meaningful planning.

The Task Force considered four general arenas within the digital repository environment: library users, the general technology environment, library collections and services, and key policy developments at the national and institutional levels.<sup>13</sup>

### *1. The user environment*

Library services are shaped to library users — present and future. By 2015 the faculty, staff, and students at research institutions will have developed new skills in the digital environment, heightened expectations of the services they use, and new sets of practices in conducting research and scholarship. Users' skill set will be grounded in their experiences of the open Web as well as their training and education in research techniques. Their research practices will have adjusted naturally as new resources and services develop and as new research questions take shape in different disciplines. As users' experience with digital content, services, and infrastructure shifts, their expectations will evolve. The directions of these changes are

**already clear — users will expect to find nearly all content in digital form and for all services to be available regardless of location.**

Most of today's junior faculty will have tenure in 2015, some of today's college seniors will be junior faculty, and today's fifth-graders will be freshmen; but regardless of their age, all faculty and students will have acquired new skills both from their research and learning environments and from their experiences of the network generally. As both content producers and consumers, they will be able to create, manipulate, and manage content in a world that is dominated by networked digital information.

Research practices will increasingly take advantage of strategies predicated on the availability of large amounts of widely accessible, rather than isolated and sparse, data. Many primary source materials supporting humanistic investigations — large corpora of texts, collections of images, and collections of cultural materials — will be complemented by many newly available and discoverable materials from disparate sources outside of library collections. To draw on content from these diverse sources, researchers will integrate use of library services and resources with funder-supported resources, commercially provided resources, and services and resources provided by other entities within the academy. Consequently, librarians will have much less control of the user experience than currently and will adopt more strategies that rely on collaboration with users. For instance, in areas such as curation and preservation of data, librarians will be regularly curating with, not just for, researchers.

Consumers of library resources will possess new tools in the form of devices, applications, and networked services. Digital anything, anywhere, anytime, and anyhow will be the expectation. The location of either the content or user will be largely unimportant. Part of this shift away from localization of content and services will be a deeper integration of access to resources and services through a variety of entry points ranging from diverse campus systems to the broader Web environment. Whether using discipline-specific simulations, analysis tools, courseware environment, archival collections on a different continent, or a popular search engine, users will expect seamless access to services and resources provided by their home institution.

## *2. The technology environment*

In the 2015 technology environment, much that is currently managed locally will be managed in a distributed manner — through collaborations within institutions or among institutions or by contracting services from the commercial sector. Local management of content and storage systems will be a waning paradigm as distributed, virtualized storage becomes the norm. The repository landscape will present a rich patchwork of repositories operating at national, disciplinary, and institutional levels, but these will become less idiosyncratic and local in their supporting infrastructure.

The same overarching trend will be reflected in interactions between systems. A wide range of systems will interoperate via one mechanism or another. Technology that is open to the development of Application Program Interfaces (APIs) will become the norm and other forms of sharable application design will have powerful advantages in achieving broad adoption. Library-focused standards for interoperation will be overshadowed by more general network standards.

Large collections of content, such as datasets, will dot the landscape. Virtualized storage technologies will

be the norm. Cloud computing will be widely implemented for a range of applications. Research will grow more reliant on the production and use of large collections of data or primary source information organized into a plethora of repositories operating at national, disciplinary, and institutional levels.

In this environment, interoperation between repositories and service technologies will be a pressing priority. A range of strategies for ensuring that information is passed effectively between repositories and between repositories and other campus systems will be in place. For instance, administrative systems, course management systems, and student information systems will routinely interoperate. As a result, library infrastructure will tend to blend with campus infrastructure as campus infrastructure becomes broadly distributed and less and less localized.

The library's technical infrastructure for content management and service delivery will also be blending into the broader information environments aimed at consumers and the general public. The ability to apply Web-enabled "external standards" will be key and take priority over "library-centric" standards. Controlling the technology environment within which library content and services are delivered will be an increasingly outmoded operating strategy.

Both increased reliance on distributed content storage strategies and increased interoperability with greater openness to technologies developed for a broader market will create new opportunities for libraries and research institutions to rely on services provided by external vendors and utilities. While this will open up new strategies for leveraging scale in the development of technology services, allowing libraries to avoid investing in in-house strategies to developing many services, issues of affordability are likely to loom. Outsourcing in a distributed environment may also be problematic for some classes of content that require tighter management, for instance because of confidentiality or copyright concerns.



### *3. The shape of library collections and services in 2015*

The balance between investing in management of unique collections and supporting widely replicated content will have shifted substantially. Similarly, libraries will have reallocated resources from supporting local collections to collectively managed collections. Network technologies and digital collections will have significantly transformed traditional emphases on local, individual, and uncoordinated strategies toward new approaches that more efficiently manage collections collaboratively. At the same time, managing unique content, not just traditional special collections but entirely new kinds of works and locally-created content, will be an important emphasis for collection and management.

As users exercise new capabilities and require new services, library services will become less "localized" within the library and within campus systems and expand into the general network environment. Library services increasingly mean machine-machine interactions and will be embeddable in a variety of non-library environments. This opens the possibility for moving library services seamlessly into various user work-flows.

During the next seven years, libraries will continue to make real progress in reshaping their collections, re-locating resources for managing content, and reconfiguring their services. New technologies will change economies of scale that have historically shaped collections and services. Libraries will have to rethink their investments, stop many historic practices, embrace new priorities, and realize savings in many areas to reinvest in new kinds of collections and services. Digital conversion of traditional collections will drive new collection management and service paradigms and collaborative strategies and outsourcing will begin to predominate over local management.

Collaboration and coordinated action in a variety of areas, both within and between institutions, will

increase, as libraries experience a sharper imperative to cultivate and expose the "aggregate library resource" in order to gain greater visibility among many competing information alternatives.

As frequently replicated materials become accessible in digital form and are collaboratively managed, a new emphasis will have developed on managing locally produced works and singular collections. An emerging service category will involve supporting collection, management, and dissemination of a diverse array of intellectual outputs including digital documents, simulations, learning objects, data, images, and performances, among others. Exposing and integrating these local outputs into the broader network environment will be a common library role.

A substantial portion of library services will be comprised of sequences of machine-to-machine exchanges and many, if not all, of those machines will be managed outside the library. A harbinger of this state of affairs is the Google search that produces data from WorldCat that is resolved to the local library (which may offer a digitized version of a book).

Service planning will be informed by a series of "failures" that range from ceased publications, failed migrations, and natural disasters. Resources for content management and service development will be tightly constrained, driving collaboration and coordination as well as resource redistribution. Explicit choices will continue to characterize decisions about which services to develop and which content to preference.

### *4. The policy environment*

Both intra- and extra-institutional policy issues will be key to repository development. Accessing and using content will be defining issues. Requirements for management and access to content that arises from funded research will shape repository content and services. Institutional policies will gradually parallel funder requirements.

Intellectual property management will continue

to present considerable challenges to a wide range of management and use activities. While it is doubtful that underlying changes in copyright law will occur, some areas may well be clarified as consensus emerges around particular practices. Consequently, institutions will have to create policies to assist faculty and researchers in managing content they create.

Asset management provides a useful rubric that can be applied to any of the outputs from the processes of research and scholarship. Those assets include, but are not limited to, formal and informal scholarly works and research data. It seems evident that in the future, research funders will have implemented a wide range of policies aimed at maximizing the value of research investments by ensuring broad access to research results. Most funding bodies will require deposit and public access to publications resulting from funded research and will require actionable data management plans.

Research institutions will play a much more active role in asset management, developing local policies that reconfigure the responsibilities for making asset management decisions between the university and the scholar/researcher, and the trend will be to have a university-level, institution-wide policy of retaining partial rights to faculty publications and other research assets.

University records will form an alternate pole on a spectrum of institutional assets. Policy systems will be required tuned to the varying character of the content managed within repositories and acknowledging various legal requirements. In addition, local technology decisions, needs for interoperability, and cost issues will interact with policy strategies.

Local policies will address questions such as who exercises copyright, the uses that can be made of works and how use may be restricted (for instance, through permissions or embargos). These kinds of questions will apply not just to text and documents, but also learning objects, software, images, and video.

For many kinds of copyrighted works, digital access will operate under a different legal regime than archiving. Some (but probably not all) of the ambiguity surrounding intellectual property rights will be resolved, at least for certain kinds of digital materials in the US, in part by resolution of the Google/publisher case. There might also be helpful federal policy development around orphan works. Whatever specific accommodations emerge, some form of resolution will have opened the way for large-scale, programmatic activities in the areas of mass digitization, access to and repurposing of research data sets, and standardization of digital curation/preservation practices. Yet, libraries will face a growing dilemma if they are expected to archive content either without providing access or where access is provided through separate systems. Resources for archival functions delivered in the absence of access functions will be hard to obtain or maintain.

#### Note

<sup>13</sup> The Horizon Analysis was done in the summer of 2008, prior to final writing of the report and thus does not attempt to fully account for recent economic developments. The task force is not in a position to predict the effects of the evolving global economic crisis, but it seems reasonable to expect that the trends encapsulated here are likely to persist although their pace may be differentially accelerated or slowed.

## Positioning Research Libraries for New Roles

---

The Task Force members agree that despite the varied funding and resource challenges faced by research institutions, the delivery of repository services is a crucial function of research libraries. However, libraries are only one stakeholder constituency. Much work is being led or contributed by IT communities, for instance. Since repository services are undergoing such a dynamic process of evolution, rather than making very specific recommendations that will date immediately, a few broader areas are highlighted below for library stakeholders to monitor and seek opportunities for participation. These are arenas where ARL and the library community should focus attention and effort to enhance research library roles in delivering digital repository services. Research libraries may not be able to lead in all of these areas, but should at least be seeking to make contributions to addressing current and emerging opportunities and concerns in each issue area.

1. *Develop a deep understanding of content users' and creators' needs to underpin the development of repository-related services.*

More work is needed to help all stakeholders understand the workflows and capabilities of researchers and scholars. Rather than developing technologies and hoping they will be usefully applied, libraries need more data, and discipline-specific data, on how a wide range of service consumers — institutions, libraries, scholars, and researchers — value services and want to use content. As services reach “upstream” into the research process,

targeted studies would support more rapid and effective development of repository-facilitated services that are demand, rather than technology, driven.

An understanding of the reward systems that motivate users is also crucial to ensure that services align with rather than conflict with them. Key reward systems include prestige-based systems for institutions, research groups, and individuals; academic promotion and tenure systems; and funding allocation systems. Ultimately, repository services must act to maintain and enhance the integrity of the scholarly process. Understanding and engaging with the user environment is key to building a successful and sustainable system of services.

2. *Apply a life-cycle management framework to guide development and evaluation of services and policies.*

In managing print publications, libraries, publishers, and other contributors have a well-developed understanding of their roles, and well-developed services to fulfill them. In the emerging world of digital content, relatively little of the life cycle of digital works, objects, and data is currently well understood. Rather than trying to replicate libraries' traditional role in developing services, there is a need to step back, consider the entire life cycle of content, redefine roles for the library in managing content, and identify appropriate partners and relationships that are needed to manage the life cycle end to end.

Within a life-cycle framework, repository content is not simply another format, rather it represents a

wide variety of formats. Developing repository services requires a nuanced understanding of the type of content that requires curation, that content's shifting value over time, and the needs of both creators and users of that content.

In applying a life-cycle framework, the selection function remains essential, but there is very little experience available to assist in developing appropriate policies. Surrounded by an abundance of digital content and insufficient resources to manage all of it, libraries need a new basis for assessing the value of different kinds of content throughout their life cycles. Acknowledging that not all content has equal value and that its value may wax and wane as it moves from creation to archiving, new models for cyclical selection functions will be appropriate in many cases.

Another facet of life-cycle management is managing versioning and creating connections between related content that may be produced at different points in the research cycle. Mechanisms for creating and assuring the integrity of content and the authority of various versions of works are crude at best. This is an area that seems ripe for imaginative applications of traditional library expertise in developing metadata, standards, and expressions of relationships.

*3. Articulate a compelling value proposition for repository-related services to justify investing resources, promote partnerships, and address sustainability concerns.*

Institutions with research missions have developed an understanding of the value proposition of traditional library services: their goals, their costs, their value to users and to the enterprise of research and scholarship. As a much wider range of digital content is created through the research process, the costs and benefits of various services, curation strategies, and policy approaches are far less clear. The costs of developing new services and managing digital content over time need to be examined

and the value that libraries specifically can offer in this emerging knowledge management environment should be articulated.<sup>14</sup>

It is not just a business case emphasizing return on investment that needs to be made, but also an argument for the value of the public goods involved and the alignment of services with institutional mission and responsibilities.

In the evolving economic environment of resource constraint and conflicting priorities for research institutions, a concise and compelling articulation of the impact repository services can achieve will become even more important.

*4. Integrate into emerging services the diverse content collections that have accumulated and will continue to arise outside of library-managed repositories.*

As creators of various content collections move on to new interests or, for other reasons, are no longer able to manage content, libraries must be prepared to take over. Some of these collections are likely to accumulate within institutional boundaries, but there is also a growing body of content that is accumulating beyond the infrastructure managed by research institutions. For instance, blogs maintained on commercially managed servers, or even e-mail hosted by providers like Google and Yahoo, will require active selection, perhaps in combination with collaborative collection strategies. Small publishers, scholarly societies, or managers of resources developed through the open Web may be at risk of "falling through the cracks." Readiness to adopt established collections of digital content is required along with mechanisms for promoting repository services.

*5. Participate actively in shaping the technology of repositories, particularly the mechanisms by which repositories make services possible.*

ARL libraries will, in many cases, be the implementers of repository technologies. As such, they are powerful representatives for many key repository users. The research library community must work collectively to define basic services and engage collaboratively in defining the functions of APIs, as well as ensuring a high degree of interoperability across different repository systems.

The task force believes that there continues to be room for diversity in the development of repository systems and that there is a strong need for diversity in the types of services made available from repositories. Nevertheless, the common interests and common experiences of research libraries should be brought to bear on defining repository systems and services so that our scarce resources can be more effectively applied in getting the maximum benefit from community software development efforts.

6. *Negotiate the significant uncertainties existing in the current rights environment and build a broader consensus about the appropriate rights environment needed to support the research enterprise in a digital environment.*

Maximizing the value of repository services requires creating a supportive rights environment that enables curation and use of content. There are substantive local, national, and international challenges both in policies and practices. Libraries must continue to advocate for the greatest possible access to content while working to ensure that they also have the rights necessary to engage in curating and archiving content.



## Call to Action to ARL Member Libraries

---

Repositories are rapidly becoming ubiquitous in research institutions and libraries need to play an active role in service development. Even where libraries are reassessing their service portfolios in response to budget reductions, each needs to be developing expertise in this arena to participate in shaping these essential services. Yet, the current volatility of the repository environment and service models makes it counterproductive to recommend particular strategies for broad adoption.

When research libraries are committed to local service development, individually and as a community, a common pool of shared experience can evolve to inform planning for repository-related services. Diverse experiences with seed collections will deepen understanding of user needs, encourage experimentation with different organizational frameworks, test different business models, and clarify options for managing technical infrastructure, making it possible to promulgate best practices and tackle system-wide issues.

Important actions that research libraries should undertake include the following:

1. *Build a range of new kinds of partnerships and alliances, both within institutions and between institutions.*

Approaching repository development with a broader conception of service development requires libraries to leverage existing relationships and explore and develop a new range of alliances. Campus computing and information technology services, research service providers, and

other local partners will need to be engaged. In many cases it may be useful to consider extra-institutional alliances, especially ones leveraging existing partnerships that can provide a foundation for a governance structure. As virtual organizations form, consider parallel relationships to support repository service development.

2. *Base service development strategies on substantive assessment of local needs rather than blindly replicating work done at another institution.*

Libraries need to look across a range of local users, content, and potential partners as they develop services. Service development will be more successful and sustainable if it is aligned to key needs and institutional priorities. Partnerships and cooperative strategies will be most effective when they are grounded in a solid understanding of local concerns and opportunities.

3. *Engage with key local policy issues and stakeholders to encourage institutional engagement with national and international policy issues.*

Policies on author rights, deposit requirements for student-authored works like theses and dissertations, treatment of content that is licensed, and ensuring compliance with grant requirements are just a few examples of the kinds of policy issues that arise in the course of crafting repository-related services. Within individual

institutions and collectively across the academy, issues like these involve a diverse set of stakeholders and require a deliberate and organized strategy for relationship building. Libraries need to engage their campus policy makers around institutional policies that affect repository services.

It is also important to encourage institutional leaders to support key national and international policy directions on issues, such as copyright or public access to various sorts of content, that are in flux. New opportunities arising from digital, networked communication are bringing new expectations and any policy shifts will inevitably present compliance issues for individual research institutions.

*4. Develop outreach and marketing strategies that assist "early adopters" of repositories to connect with the developing repository-related service system.*

Most research institutions already house diverse collections of digital content. Especially where researchers and scholars have collected content, they may be approaching the limits of their abilities to effectively manage and curate what they have gathered. Where these collections

are of high value, local processes are needed to migrate early digital collections into an institutionally-managed service environment.

*5. Define a scope of responsibility to guide the development of repository services for varied forms of content.*

Within a life-cycle management framework, each library needs to identify the scope of responsibility it can assume for specific kinds of content and where others will be responsible for service provision. Typical issues include deciding the locus of responsibility for curation, platform development, storage management, and so on. The scope of responsibility may well vary with different kinds of repository content. Data sets, learning objects, licensed content, and university records are all examples of different content types for which libraries might assume different levels of responsibility and provide different kinds of services.

Note

<sup>14</sup> For instance, the work of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access will address some parts of this area. See <http://brtf.sdsc.edu/>.



## Conclusion

---

Digital repository services present a new frontier for research libraries to explore. As a frontier, it possesses undiscovered opportunities, surprises, and challenges. Most research libraries have begun the process of pioneering and are finding that, as with other frontiers, success requires a willingness to try new things and change course to avoid obstacles or seize new opportunities. Successful enterprises will benefit from a close observation of the landscape, making choices based on their local situation and the results of early exploration, while keeping an eye on what others are learning. Partnership, collaboration, and the creation of shared resources are key strategies.

Some may wonder if libraries can afford to develop repository services, especially in a time when research institutions face shrinking resource bases. The Task Force members believe that neither research libraries, nor the institutions they serve, can afford to do without repository services. Such services have a powerful potential to enable key work and enhance the effectiveness of a wide range of functions across research institutions. Researchers and scholars with access to a spectrum of

repository services will possess a substantial advantage in conducting cutting edge research, delivering high quality teaching, and contributing valuable services to society.

The Task Force members hope that this report presents a useful perspective on the digital repository environment and inspires ARL member libraries and others to assess their views and plans for service development. Repository management will not be the sole purview of libraries, by any stretch of the imagination, but libraries have key strengths and, arguably, missions requiring them to undertake various roles in repository service development.

Finally, they urge ARL to incorporate the report's findings into its program planning and engage in the issue arenas the task force has identified. On an ongoing basis, the organization should measure or monitor ARL libraries' repository service development and act to support members in their service development. ARL is well-positioned to act to help members act in coordination with regard to key policy issues and share best practices.



## Selected Bibliography

---

- "Agenda for Developing E-Science in Research Libraries: ARL Joint Task Force on Library Support for E-Science Final Report & Recommendations," November 2007. [http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf).
- Allinson, Julie, Sebastien Francois, and Stuart Lewis. "SWORD: Simple Web-service Offering Repository Deposit." *Ariadne* (January 30, 2008). <http://www.ariadne.ac.uk/issue54/allinson-et-al/>.
- Bailey, Charles W. *Institutional Repositories*. SPEC Kit 292. Washington, DC: Association of Research Libraries, July 2006. <http://www.arl.org/bm~doc/spec292web.pdf>.
- Center for Research Libraries Case Studies. "... the case studies will identify practices, strategies and mechanisms that have enabled repositories to sustain massive digital collections over substantial periods of time."
- McCargar, Victoria. "Repository Profile: The Associated Press." October 15, 2008. [http://www.crl.edu/PDF/AP\\_Profile.pdf](http://www.crl.edu/PDF/AP_Profile.pdf).
- Waltz, Marie-Elise. "Repository Profile: NORC General Social Survey." August 19, 2008. [http://www.crl.edu/PDF/NORC\\_profile.pdf](http://www.crl.edu/PDF/NORC_profile.pdf).
- Crow, Raym. "The Case for Institutional Repositories: A SPARC Position Paper." *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, no. 223 (August 2002) <http://www.arl.org/bm~doc/instrepo.pdf>.
- Digital Academic Repositories (DARE) Project, SURF Foundation, Netherlands. <http://www.surfoundation.nl/smartsite.dws?fs=&ch=ENG&id=5377>.
- Foster, Nancy Fried, and Susan Gibbons. "Understanding faculty to improve content recruitment for institutional repositories." *D-Lib Magazine* 11, no. 1 (2005) <http://www.dlib.org/dlib/january05/foster/01foster.html>.
- Foster, Nancy Fried, and David Lindahl. 2008. "Enhancing E-Resources by Studying Users: The University of Rochester's Analysis of Faculty Perspectives on an Institutional Repository." preprint: <http://hdl.handle.net/1802/5888>.

Joint Information Systems Committee (JISC)

Jacobs, Neil. "Digital Repositories programme 2005–7." <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005.aspx>.

—. "Digital Repositories programme 2007–8." <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007.aspx>.

Thomas, Amber. "Repositories and Preservation programme." <http://www.jisc.ac.uk/whatwedo/programmes/reppres>.

Jones, C., R. Darby, L. Gilbert, and S. Lambert. "Report of the Subject and Institutional Repositories Interactions Study. Monograph." JISC, November 25, 2008. <http://ie-repository.jisc.ac.uk/259/>.

*Journal of Digital Information*. 8, no. 2 (2007); Special issue: "Digital Curation and Trusted Repositories." <http://journals.tdl.org/jodi/issue/view/34/showToc>.

Lavoie, Brian, Lorraine Eakin, Amy Friedlander, Francine Berman, Paul Courant, Clifford Lynch, and Daniel Rubinfeld. 2008. "Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation. Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access." [http://brtf.sdsc.edu/biblio/BRTF\\_Interim\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf).

Lynch, Clifford A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, no. 226 (February 2003) <http://www.arl.org/bm~doc/br226ir.pdf>.

Lynch, Clifford A., and Joan K. Lippincott. "Institutional repository deployment in the United States as of early 2005." *D-Lib Magazine* 11, no. 9 (2005) <http://www.dlib.org/dlib/september05/lynch/09lynch.html>.

Markey, Karen, Soo Young Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Yakel. *CLIR Report-pub 140*. CLIR Reports. Washington, DC: Council on Library and Information Resources, February 2007. <http://www.clir.org/pubs/abstract/pub140abst.html>.

OCLC, Center for Research Libraries, National Archives and Records Administration. "Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist." <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>.

Palmer, Carole L., Lauren C. Tefteau, and Mark P. Newton. "Identifying Factors of Success in CIC Institutional Repository Development - Final Report." The Andrew W. Mellon Foundation, August 2008. <http://hdl.handle.net/2142/8981>.

Poynder, Richard. "Clear Blue Water. Blog." EPrints Insiders. March 1, 2006. <http://www.eprints.org/community/blog/index.php?/archives/43-Clear-blue-water.html>.

Randall, Ryan, Jane Smith, Katie Clark, and Nancy Fried Foster. "The Next Generation of Academics: A Report on a Study Conducted at the University of Rochester." *Scientific Commons* (2008 ) <http://hdl.handle.net/1802/6053>.

Rieger, Oya Y. "Opening Up Institutional Repositories: Social Construction of Innovation in Scholarly Communication." *Journal of Electronic Publishing* 11, no. 3 (Fall 2008) <http://dx.doi.org/10.3998/3336451.0011.301>.

Research Libraries Group. "An Audit Checklist for the Certification of Trusted Digital Repositories." Mountain View, CA: RLG, 2005. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>.

University of California Office of Scholarly Communication and California Digital Library eScholarship Program. "Faculty Attitudes and Behaviors Regarding Scholarly Communication: Survey Findings from the University of California." Berkeley, CA: University of California, 2007. <http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf>.

van Westrienen, Gerard, and Clifford A. Lynch. "Academic Institutional Repositories: Deployment Status in 18 Nations as of Mid 2005." *D-Lib Magazine* 11, no. 9 (2005) <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>.



## Appendix

---

### *Summary of responses from other organizations regarding activities around digital repositories*

#### **Canadian Association of Research Libraries (CARL)**

Report prepared by Kathleen Shearer on behalf of CARL.

There are two programs at CARL related to digital repositories:

#### **1. The CARL Institutional Repository Program**

The Canadian Association of Research Libraries (CARL) began the Institutional Repository Project in 2003 in order to provide support for Canadian implementers of institutional repositories (IRs). The project conducts an annual survey of IRs in Canada and maintains a listserv and to date has completed the following activities:

- Development of the CARLCore Metadata Application Profile
- Development of a Digital Repository Brochure (with SPARC)
- Implementation of a CARL Harvester to aggregate Canadian IR content
- Work with Creative Commons Canada to adapt the Creative Commons License to the Canadian environment
- Organizes meetings and conferences on institutional repositories

The program is guided by a Working Group composed of several CARL Directors: Carol Hixson, University of Regina(chair); Pam Bjornson, CISTI – NRC; Lucie Gardner, Université du Québec à Montréal; Thomas Hickerson, University of Calgary; Paul Wiens, Queen's University, and CARL Research Associate, Kathleen Shearer.

The Terms of Reference for the Working Group are as follows:

The Working Group will develop a vision for institutional repositories in Canada; set priorities and specific projects for the CARL Institutional Repository Program; and, coordinate the implementation of projects.

The Working Group will meet regularly by teleconference and at CARL Meetings to review its vision, priorities, and projects.

The Working Group will report on Program activities to CARL members and ensure that the membership has input into the development of priorities and activities.

The Priorities for the CARL Institutional Repositories Program are:

To articulate the importance of institutional repositories and the value of the CARL IR Program.

To improve content recruitment at institutional repositories in Canada.  
To demonstrate the value of overlay services for Canadian institutional repositories.

### *Current Projects*

#### Coordinated Communication Strategy

The CARL IR Working Group has been working with Canadian Institutes of Health Research (CIHR) to develop a coordinated communication strategy for the "CIHR Policy on Access to Research Outputs." The success of the policy rests to a large extent on awareness of the policy on campus. Researchers need to know that the policy exists and also how to comply with it (especially in terms of understanding publisher self-archiving policies). CARL's network of libraries across Canada could be instrumental in assisting with the implementation of the CIHR policy. The strategy proposes to undertake series of presentations across Canada that will introduce the CIHR access policy to researchers on campus.

#### Interoperability between IRs and disciplinary repositories

CISTI and CARL are conducting a feasibility study on the implementation of the SWORD protocol to enable simultaneous deposit into multiple repositories. The study will survey CARL libraries to determine interest in implementing SWORD. This will be followed by a Pilot Project to implement the protocol at interested repositories.

#### Usage Statistics

Usage statistics are a very powerful tool for promoting and populating repositories. This project will review existing methods for collecting usage statistics in Canada and elsewhere and provide a document outlining best practices for IRs in Canada.

### *Future Projects*

#### Single-disciplinary Pilot Project

Over the last several years, many services have been built into repository software platforms (such as hit counts on papers, search engines, and personalized publications lists, etc.). However, we have not yet seen a lot of services built on top of the aggregate content of IRs. The hypothesis is that researchers are much more likely to deposit their research output if they are contributing to a disciplinary-based collection, rather than institution-based. This project will harvest material in a single-discipline across several repositories and build value-added services on top of the content. The project will involve four or five institutions and will assist in determining the projected operating costs for this type of service. The Project will encourage the participation of granting councils and faculty associations and possibly seek external funding.

#### CARLCore Metadata Profile: Phase Two

Many of the functionalities of the new types of overlay services discussed above require more comprehensive metadata than is currently being assigned at Canadian institutional repositories. While it is foreseeable that certain metadata elements can be assigned after content has been archived (for example, through social tagging of subject headings), this is not feasible for some types of metadata elements such as sponsoring agency, discipline, or peer-review. This project will further develop the CARLCore Metadata Profile to include the necessary metadata elements that enable the development of overlay services for harvested content.



## 2. The CARL Data Management Working Group

In the spring of 2007, CARL conducted a survey of member activities in regards to researcher-generated data. The survey found that most member libraries are interested in storing this data, but few have a formal data archiving policies or appropriate infrastructure.

To address this, CARL formed a Data Management Working Group in May 2007. Members of the Working Group are: Marnie Swanson, University of Victoria (Chair); Pam Bjornson, CISTI – NRC; Lynn Copeland, Simon Fraser University; Michelle Edwards, University of Guelph; Margaret Haines, Carleton University; Janine Schmidt, McGill University; Kathleen Shearer, CARL Research Associate; Diego Argáez. Research Officer  
Canadian Association of Research Libraries

Much of today's research involves creating and analyzing vast amounts of data of unprecedented size and complexity; and storing it for possible future use. In the spring of 2007, CARL conducted a survey of member activities in regards to researcher-generated data. The survey found that most libraries are interested in storing this data, but few have a formal data archiving policy. To this end, CARL is forming a Data Management Committee to assist members in collecting, organizing, preserving and providing access to the research data generated at their own institutions and to formulate a cooperative approach for CARL.

### Terms of Reference

- To undertake a survey of international initiatives related to the management of researcher-generated data.
- To undertake a survey of existing Canadian initiatives related to the management of researcher-generated data.
- To identify options for the role for libraries in collecting, organizing and providing access to researcher-generated data.
- To develop "best practice" models or frameworks for the management of researcher-generated data to be used by individual CARL members.
- To develop a plan for a cooperative approach to be adopted by CARL in relation to the collection and organization of research-generated data (i.e., because data management practices are very discipline specific; one possible model is "centers of excellence" for different disciplines at different institutions).

### *Current Activities*

- Develop a Data Management Awareness Toolkit for CARL directors to enable them to raise awareness of data management issues on campus.
- Identify appropriate roles for libraries in the area of data management.
- Design a strategy for developing expertise in data management in libraries.
- Create a program on Data Management for an upcoming CARL meeting.

## **Council on Library and Information Resources (CLIR)**

Provided by Amy Friedlander, CLIR

CLIR does not maintain a digital repository. More generally, we are engaged in a range of initiatives and projects that bear upon the shape and form of such systems. These include:

- Two reports, one on institutional repositories and a second on e-journals. Kathlin wrote a story for the special issue of the *Journal of Electronic Publishing* (January 2008), in which she reviewed these reports together and pointed out their resonances.
- A recent paper by Dawn Schmitz where she looked at the literature on institutional repositories and cyberinfrastructure.
- Support for and participation in the NSF Blue Ribbon Task for on Economically Sustainable Digital Preservation and Access.

Our research programs in cyberinfrastructure, preservation, and digital scholarship all have implications for repository systems but do not take on the issue directly. So again, in a general sense, our programs may affect the discussions (for example, by considering the publication status of contributions to disciplinary repositories such as the PDB and the implications for promotion and tenure and incenting contributions to archiving and preservations systems) but we do not presently have research that directly addresses repositories.

## **NISO: National Information Standards Organization**

Prepared by Todd Carpenter

### *Ongoing programs*

NISO has no ongoing programs explicitly committed to digital repository issues, however it seems likely that at any given time some of its standards work will be related to digital repositories.

### *Special initiatives/projects*

- Planned: NISO received Mellon funding to hold a Thought Leader Meeting on Institutional Repositories. A strong recommendation was to work on solutions to the barriers authors face in placing their work into multiple repositories. Actions to promote the development of a common deposit mechanism with metadata standards were recommended.
- JAV (Journal Article Versions) Working Group  
The work plan will include:
  1. Creation of use cases to identify the most common journal article life cycles.
  2. Analysis of use cases to determine common life cycle stages.
  3. Selection of preferred vocabulary for the most common life cycle stages.
  4. Development of appropriate metadata to identify each variant version and its relationship to other versions, in particular the definitive, fully functional published version.
  5. Establishment of practical systems for ensuring that the metadata is applied by authors or repository managers and publishers.

## **SPARC & SPARC Europe**

Provided by Jennifer McLennan

### *Ongoing programs*

- SPARC, SPARC Europe, and SPARC Japan together host regular meetings on Digital Repositories. The meetings focus on policy and advocacy issues over technology issues. The last meeting took place in Baltimore on November 17–18, 2008.
- SPARC and SPARC Europe partially fund the Directory of Open Access Journals.
- SPARC Europe partially funds OpenDOAR the Directory of Open Access Repositories.
- Repositories are a focus of the SPARC Japan agenda as well. Their focus is on “enhancing institutional repository friendly policy choices” and “recommending open access with the author-pay model.” SPARC Japan recently forged a partnership with DRIVER, which supports large-scale collaboration on a repository collective.
- SPARC is the host of the SPARC-IR discussion list.

### *Forthcoming events*

- The SPARC Digital Repositories Meeting 2010  
Due to the tremendous success of the 2008 meeting, and demand from the community, SPARC will host the next North American repositories meeting in November 2010.
- Web casts  
In the interim, SPARC will host a series of Web casts that will examine in more detail topics introduced at the in-person meeting.
- Digital repositories coalition  
Also in follow up to the November meeting, SPARC will explore and support the formation of a coalition of repository advocates, similar to the Open Access Working Group or OASPA, which will collaborate on initiatives related to education and advocacy.
- Greater Reach for Your Research - Repositories brochure and educational campaign  
In partnership with the Canadian Association of Research Libraries, SPARC will introduce a new initiative to engage Canadian researchers on the topic of digital repositories. Scheduled for launch in early 2009, the campaign will include a color brochure to be distributed through CARL member libraries, a video slide show version of the message, and other Web elements. A focus of the campaign will be encouraging CIHR-funded researchers to deposit manuscripts in the library repositories.  
The SPARC Steering Committee has also expressed support for an American version of this campaign to be released.

### *Special initiatives*

- SPARC's ongoing advocacy for public access through the National Institutes of Health, other US agencies, the Canadian Institutes for Health Research, and other emerging policies, supports the development of agency- and institutionally based digital repositories.

