

Leveraging Short-term Opportunities to Address Long-term Obligations: A Perspective on Institutional Repositories and Digital Preservation Programs

NANCY Y. MCGOVERN AND APRILLE C. MCKAY

ABSTRACT

Until now, information retrieval (IR) managers have been absorbed by efforts to increase the amount and quality of scholarly deposits. Other pressing concerns have been to develop the software, standards, and other tools to insure access, exchange, and discovery of the works in the IRs. But building an IR without making plans for technological, organizational, and resourcing sustainability is like building a house on sand.

At this particular juncture, there are opportunities to enhance the efforts of both institutional repository implementation and digital preservation program development by bringing together the strengths of each. This paper first explores the developmental paths and intersections of digital preservation and institutional repositories, considers the current status of both, and looks ahead toward the opportunities and challenges inherent in their convergent future.

INTRODUCTION

In pursuing the compelling goals of public access to scholarly output and collective economic action against high subscription prices, the open access community has mobilized organizational support and resources to create important infrastructure to support the exchange of scholarly content, and free it from the sometimes overreaching grasp of journal publishers. From a preservation perspective, however, there are serious questions about whether sponsoring organizations can sustain their commitment to archive the digital content deposited in institutional repositories. At this particular juncture, there are opportunities to enhance the efforts of both institutional repository implementation and digital preservation program development by bringing together the strengths of each.

This paper first explores the developmental paths and intersections of digital preservation and institutional repositories, considers the current status of both, and then looks ahead toward the opportunities and challenges inherent in their convergent future.

The movement promoting access to publicly funded research has enjoyed snowballing success over the last several months. In a signal legislative victory, the National Institutes of Health now requires researchers it funds to deposit copies of their peer-reviewed manuscripts in PubMed Central upon their acceptance by a journal (NIH, 2008). Journals' need for commercial exploitation of the works has been protected by a one-year embargo on the public release of a free version of the works; but the law has helped organize authors to make more favorable intellectual property licenses with publishers. The other chief source of federal science research dollars, the National Science Foundation, has embraced a vision of public access to scientific data, and plans invest in "cyberinfrastructure" to enable the deposit and reuse of data (NSF, 2007). Pressing the NSF to commit to archiving of *published* research results is currently a focus of the open access community.

Other recent open access successes include resolutions of the Harvard Arts and Sciences and Law faculties in the spring of 2008 to require faculty members to allow the university to provide free access to their work online (Guterman, 2008). A similar mandate from the high-energy physics community has produced the SCOAP³ initiative, a consortium of laboratories and universities that have pledged to redirect the funds formerly used to subscribe to physics journals to help convert the field's journals to open access. (SCOAP³, 2008). All of these efforts to support open access raise pressure on institutions to create institutional repositories (IRs) and place even more urgency on the issue of long-term sustainability of content.

Until now, IR managers have been absorbed by efforts to increase the amount and quality of scholarly deposits. Other pressing concerns have been to develop the software, standards, and other tools to insure access, exchange, and discovery of the works in the IRs. But building an IR without making plans for technological, organizational, and resource allocation sustainability is like building a house on sand.

WHERE WE CAME FROM

This section compares the key characteristics and developmental milestones of institutional repositories and digital preservation programs as background for the discussion. It briefly discusses digital preservation community standards and practice in relation to institutional repositories.

Digital Preservation Milestones

Through the 1980s and 1990s the advent of personal computers and new technologies such as the Internet and e-mail transformed human com-

munication and recordkeeping practice. Greater access to computers by more diverse users created an explosion in the amount and variety of digital content. Though there were earlier precedents for preserving digital content dating from the 1960s, the emergence of the digital preservation community can be dated from 1996.

In December 1994, recognizing that archivists and librarians had a responsibility to learn how to keep digital materials accessible, the Commission on Preservation and Access and the Research Libraries Group (RLG) created the Task Force on Digital Archiving. The task force, with members drawn from archives, libraries, publishers, scholarly societies, government, and business, issued its final report in 1996, entitled "Preserving Digital Information" (Waters & Garrett, 1996). The work has proven influential, and has helped to define a research agenda for digital preservation for more than a decade. It identified the need for deep infrastructure to support digital archiving, and mapped specific strategic research goals for building it. The cochairs, Don Waters and John Garrett outlined the import of the work:

If we are effectively to preserve for future generations the portion of this rapidly expanding corpus of information in digital form that represents our cultural record, we need to understand the costs of doing so and we need to commit ourselves technically, legally, economically and organizationally to the full dimensions of the task. (Waters & Garrett, 1996, 3-4)

The development of the Open Archive Information System (OAIS) by the Consultative Committee for Space Data Systems (CCSDS) of NASA, beginning in 1995 was a key step in the construction of the standards infrastructure for digital preservation. (CCSDS, 2002) The ISO standard includes terminology and concepts for describing and comparing archival architectures and operations. It also propounds a "roadmap for the development of related standards," which includes mechanisms and methods for archival interface, ingest, delivery, identification, search, and retrieval, and a call for a standard for the accreditation of archives.

Efforts to develop standards in several of these areas were already well underway, progressing in tandem with OAIS. As the OAIS standard was maturing and gaining acceptance, the Digital Archive Directions (DADs) workshop, held in 1998, targeted the crucial areas of ingest, identification, and certification of archives for attention. Taking up the DADs charge, the Archival Workshop on Ingest, Identification, and Certification Standards (AWIICS) in 1999 built an agenda for the creation of standards to describe and certify digital archives.

These efforts resulted in several important projects. First, RLG and OCLC pushed to make progress on one of the goals originally set in Waters and Garrett's "Preserving Digital Information," report—to create a definition of a trusted digital repository (TDR) and to outline expectations for

institutions that aimed to preserve digital cultural resources. (RLG-OCLC, 2002). The TDR enumerates characteristics of a sustainable digital repository for large-scale heterogeneous research collections, including:

- OAIS compliance
- administrative responsibility
- organizational viability
- financial sustainability
- technological and procedural suitability
- system security, and
- procedural accountability.

In addition, the report discussed methods and strategies for certification of TDRs, so that stakeholders (depositors, researchers, funders, etc.) would not need to take a repository's self-declaration of trustworthiness at face value. Five years later, *Trusted Digital Repositories Audit and Certification* (TRAC), by the RLG-NARA Digital Repository Certification Task Force (2007), and supported by the Center for Research Libraries (CRL), delineated a set of metrics against which to measure progress toward "trusted repository" status.

Meanwhile, in Europe, the Digital Curation Centre (DCC) and Digital-PreservationEurope (DPE) were developing a tool to identify and manage the risks and uncertainties associated with digital preservation using a self-assessment model. The DRAMBORA toolkit, released in draft in February 2007, guides repository managers through a rigorous self-audit. It helps to document organizational commitment to preservation, the policy and regulatory framework and work processes surrounding preservation activities, and to identify and manage risks to digital content.

The German collaborations, nestor and DINI, which were also working on the problem of repository certification and audit, emphasized the importance of coaching repositories toward good practice; providing tiered certifications so that the bar is not set so high for young organizations that their participation is discouraged. (nestor, 2006; DINI, 2003).

Realizing that they needed to coordinate their energies and resources, the leaders of four digital preservation organizations (DCC, DPE, nestor, and CRL) convened in Chicago in January of 2007 and crafted "Core Requirements for Digital Archives." Since then, a new project has carried on work to create an ISO standard against which a full audit and certification of digital repositories can be based. Under the standards development auspices of the CCSDS (which sponsored OAIS), this group aims to create a standard that will allow self-assessment as well as external audit, and that will provide the basis for tool development and best practice guides (Digital Repository Audit, 2008). The standard uses a risk-assessment approach, rather than propounding mandates, so that different repositories can elect the best policies and strategies for their particular circumstances.

In addition, the drafters recognize that best practices are constantly changing, and therefore intend that the standard will use a “continuous quality improvement” model so it can be flexible enough to accommodate changing demands and expectations (Digital Repository Audit, 2008).

The PREMIS working group has addressed another important standards piece of the digital preservation infrastructure—defining the metadata necessary to support long-term preservation of digital materials. PREMIS (PREservation Metadata: Implementations Strategies), jointly sponsored by OCLC and RLG, released its 237-page *Data Dictionary for Preservation Metadata* in May 2005. Version 2.0 of the dictionary was released in April 2008 (PREMIS Editorial Committee, 2008). Take-up of this standard has been limited; as of June 2008, only nine repositories are listed on the implementation registry, though the listed participants include leading organizations, such as the Library of Congress, Cornell, Oxford, Stanford, and the National Archives of Scotland.

Institutional Repositories Milestones

Institutional repositories (IRs) originally emerged as an open access infrastructure to help universities combat journal publishers’ skyrocketing subscription prices and to fulfil a vision of free access to scholarly information. The Budapest Open Access Initiative (BOAI), the product of a meeting of the Open Society Institute in December 2001, provided marching orders for librarians and academics who dreamed of making research articles in all academic fields available for free online. Tools and assistance were needed, it said, to help scholars to self-archive their work in open electronic archives (Budapest, 2002). Conceived as a way for universities to capture, preserve, and provide free access to their members’ intellectual output, IRs were deployed primarily in academic settings (Ferreira, Rodriguez, Baptista, & Saraiva, 2008).

Before BOAI, the Los Alamos Physics Archive, now known as the arXiv, served as a *discipline-based* self-archiving depository, which by 1999 had accumulated over 100,000 papers deposited by their authors. In 2001, Cornell University assumed managerial responsibilities for arXiv where it flourishes, with more than 100,000 distinct users per day. Other smaller discipline-based archives included CoRR in computer science, CogPrints in the cognitive sciences, and PubMed Central (Harnad, 1999). The idea of *institution-based* electronic repositories had been advanced as early as 1994, but had not had much success until software tools and metadata standards began to emerge over the next decade (Okerson & O’Donnell, 1995).

Notably, until the Open Archives Initiative’s Protocol for Metadata Harvesting (OAI-PMH) limned out a framework in 1999 for interactive sharing among independent sites, IRs could not compete with the discovery capabilities afforded by topical repositories (Hitchcock, Brody, Hey, & Carr, 2007).

With the release of the first version of DSpace (<http://www.dspace.org/>), in November 2002, many universities could begin to contemplate the creation of an IR. The open-source software package, codeveloped by MIT and Hewlett-Packard, provides tools for managing digital assets. As of May 20, 2008, there have been 324 installations of DSpace in 54 countries (DSpace, 2008). Other software packages, such as Eprints (<http://www.eprints.org/>) (the first IR software), bepress (<http://www.bepres.com/ir/>) and Fedora (<http://www.fedora-commons.org/>), provide alternative software choices for IR installations.

With such a strong vision of the value of open access to scholarly output, IR builders were surprised and disappointed by the low participation by faculty members. Several studies measured and described the growing body of IRs by examining the character of the depositors, and the type of material deposited (Lynch, 2003; Davis & Connolly, 2007; Rieh, Markey, St. Jean, Yakei, & Kim, 2007; McDowell, 2007; Thomas & McDonald, 2007). The results were disheartening. McDowell (2007), for example, found that for all active IRs over a twelve-month period, from November 2005 to November 2006, the average growth in the number of IR deposits was 1,100 and the median was only 366 items, or one a day. Furthermore, she determined that fully 41.5 percent of the material deposited was student work, particularly electronic theses. Faculty deposited only 37 percent of the objects, and only 13 percent of the whole consisted of peer-reviewed work (McDowell, 2007).

Since many of the IRs had been specifically created to provide a place for professors' self-archived peer-reviewed work, IR managers struggled to provide strategic plans to achieve a higher rate of these deposits. They have devoted energy to creating marketing plans, negotiating deposit mandates with university administrators, and providing cash incentives for deposits. (Ferreira, Rodriguez, Baptista, & Saraiva, 2008). They have created and shared tools to help with deployment, recruitment, and marketing. (Estlund & Neatrou, 2007). They have persuaded scholarly publishers to allow self-archiving and have compiled registries listing publishers' licensing policies (SHERPA, 2008).

In short, IR managers have been so distracted by access and ingest issues that very little attention has been given, to date, to the problem of how promises to preserve these materials will be honored.

WHERE WE ARE

There is no question that institutional repositories offer opportunities as well as challenges for digital preservation. In this section we identify them.

Institutional Repositories Opportunities for Digital Preservation

Institutional repositories offer at least five significant opportunities for digital preservation.

First, and perhaps most importantly, institutional repositories focus organizational attention on managing digital content. Through effective negotiation and lobbying, IR advocates have parlayed their passion for open access into real influence. Their articulation of a vision for the free exchange of information has proven persuasive to decision makers and resource granters. By convincing these players that these new assets must be protected by sound policy and preservation practice, their attention could be leveraged for achieving digital preservation goals.

Second, institutional repositories provide a potential entry point—even a back door—for getting content into digital preservation programs. Since the material that is deposited is often selected by its creator, there are some types of content that will end up in IRs that would not normally be available for appraisal by a conventional archive. Evidence of teaching and student life, for example, is often less well-documented than university archivists would wish (Prom & Swain, 2007). Gray literature, such as technical reports, conference proceedings, white papers, or reports of working groups and committees, is another example of scholarly output that might not otherwise be preserved and accessible in the long term.

Third, depositors and other stakeholders in institutional repositories may learn about digital preservation issues when they deposit digital content into IRs. For example, online FAQ's and submission instructions for IRs typically include a discussion of which file formats will be supported and whether the functionality of a document will be preserved, or merely its intellectual content. By participating in the ingest process, depositors may come to understand what kinds of metadata and formats can aid the long-term management of digital materials, and this knowledge may influence their personal recordkeeping practice.

Fourth, IRs may offer an opportunity to address preservation planning priorities by providing guidelines and tools for depositors to prepare archive-ready digital content. Since IRs have worked so hard on education and outreach to depositors, they may be perfect laboratories for testing and improvement of templates and submission techniques to improve the quality of submissions.

Fifth, as faculty members near retirement, their desire to preserve their digital legacy is a strong incentive to deposit significant materials in institutional repositories. These materials may present preservation challenges, because they may be in old formats or contain poor or nonexistent metadata. Solving the problems of managing end-of-career bulk submissions could further the goals of both IRs and digital preservation programs.

Institutional Repositories Challenges for Digital Preservation

Though progress has been made in addressing preservation concerns, some significant challenges remain. There is typically little control over what is ingested into an institutional repository. Allowing depositors to

choose the content and the format of submissions is a perceived incentive for self-archivers. But managing these deposits responsibly is a challenge for digital preservation because digital content that is difficult or costly to preserve or that has no known digital preservation strategy may be ingested. Since the managers do not assess the value of the information to be preserved, it is difficult to make good judgments about how much preservation investment is warranted.

Second, IRs' difficulties creating a critical mass of deposits have led them to try to make the deposit process as easy as possible for busy professors. This can result in the deposit of materials in less-optimal formats, with poor metadata and insufficient intellectual property rights clearance. Good digital preservation practice recognizes the need to balance producer and archive responsibilities so that managers are not saddled with the responsibility of preserving content that is expensive to maintain, or impossible to provide access to.

In the same vein, there may be no recourse for institutional repositories if producers deposit malformed Submission Information Packages (SIPs). Under OAIS, a SIP is the combination of the deposited digital object and the information associated with its submission by the producer. One of the attributes of a trusted digital repository is that it is subject to audit and report procedures to insure that the SIPs are valid, so that the repository can timely cure errors. Typically these procedures would include a renegotiation of the transfer with the producer if a SIP was malformed or corrupted. But IRs may have policies or practices in place that prevent the administrator of the repository from contacting the depositor to fix the bad submission. Uncured, these errors can lead to the preservation of unreadable data and disappointed user expectations.

Slower-than-expected take-up for institutional repositories has focused a disproportionate amount of collective effort on ingesting materials. Simplifying deposit and rights clearance, negotiating institutional deposit mandates, and creating outreach and marketing plans are all important activities to insure the success and sustainability of IRs. But sustainability efforts should be spread out across the system functions and not clustered entirely at the ingest end. Unless plans and policies for good data stewardship support the management of these assets, technological, organizational, and resource changes can place their survival at hazard. Soliciting materials for deposit without devoting resources to planning for their preservation or explicitly stating that preservation will not be undertaken is irresponsible. Feedback from participants in the Digital Preservation Management Workshops has indicated that IR managers do understand and accept this, but may have difficulty diverting resources from ingest activities toward preservation.

Finally, and possibly most endemic, preservation is often simply overlooked since access is the driver for institutional repositories. But preser-

vation is access on a longer timescale, and we cannot expect producers to continue to deposit scholarly materials in our institutional repositories if the commitment to keep them accessible in the long term is ambivalent and ambiguous. IRs have undertaken to provide *better* access to scholars' work than publishers do—better can, of course, mean cheaper and faster—but it can and should also mean longer.

Many IRs have no explicit mandate for digital preservation. Depositors may expect or assume long-term preservation when they deposit—an expectation perhaps fostered by an IR at the depositor's last employing university. In an era where the preservation community is defining the attributes of trusted digital repositories, IRs inject a measure of uncertainty and variability that may be confusing for users. Solving the challenges enumerated above will create strengthened IRs and digital preservation programs at host institutions, and promote the emergence of best practices and standards that can inform good digital curation in many contexts. Viewing IRs as so many laboratories for research and development of preservation tools and expertise allows us to imagine a world in which managing IRs foster many valuable competencies.

WHERE WE ARE GOING

In the next section, we discuss the organization context for digital preservation and institutional repositories, and present our five-stage model for the development of a mature program for the preservation of digital objects in repositories.

Organizational Approaches

To help managers overcome technological, resourcing, and organizational impediments to setting up secure preservation programs, Anne R. Kenney and Nancy Y. McGovern developed a five-day Digital Preservation Management Workshop and online tutorial (ICPSR, 2007.) at Cornell with funding from the National Endowment for the Humanities. As of 2008, the workshop is hosted at the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR). Developing and teaching the workshop has taught us about the stages organizations pass through when grappling with new missions and responsibilities. One of the most significant things learned is that amid the distractions of technological change, insufficient attention was being paid to the organizational context of digital preservation programs (Kenney & Buckley, 2005). Organizations cannot acquire ready-made, out-of-the-box digital preservation programs. Rather, every program is uniquely situated within its institutional context, and defined and constrained by the particular objects to be preserved and the existing technological infrastructure. We came to understand that cultural repositories pass through five stages on their way to developing a fully mature digital preservation program. Each

stage is clearly delineated and is characterized by key attributes and organizational responses. Our conclusion was that organizational readiness, not technology, was the greatest inhibitor for building sustainable digital preservation programs (Kenney & Buckley, 2005).

The organizational approaches to the management of institutional repositories and digital preservation follow a similar developmental pattern. In the three observable models we perceive a familiar progression:

- *Segregated*: the management of institutional repositories and digital preservation programs as distinct programs or silos. IRs may explicitly rule out digital preservation responsibility (*Stage 1 & 2*)
- *Modular*: the institutional repository as the Producer (pre-ingest) and Consumer (access delivery) ends of the OAIS model (*Stage 2 & 3*)
- *Integrated*: the interconnection of institutional repositories and digital preservation programs (*Stage 4 & 5*)

In the section that follows, we outline the five organizational stages (acknowledge, act, consolidate, institutionalize, and externalize) as they relate to the development of digital preservation within institutional repository programs. The five stages enable organizations to define objectives, measure progress, and communicate outcomes.

Five Stages

In the Digital Preservation Management Workshop, we refer to a model of a three-legged stool to represent a secure foundation for digital preservation. The three legs are organization, technology, and resources; and without any one of them, the stool tips and the program falls (Kenney & McGovern, 2003). For organizations that have both institutional repositories and a commitment to digital preservation, these are the characteristics of the five developmental stages for digital preservation applied to institutional repositories using three legs of digital preservation stool:

Stage 1 Acknowledge. At Stage 1 the institution exhibits understanding that the programs are of local concern. No longer is there a sense that “it’s someone else’s responsibility” to preserve digital objects or to create institutional repositories, or that the problems will take care of themselves. Nevertheless, policies are nonexistent, implicit, or very high-level and generalized. Technical infrastructure is nonexistent, heterogeneous, decentralized, and opportunistic. Finally, the focus on the materials to be archived will be reactive, rather than encompassing the potential scope of materials that need to be preserved. Conversely, there may be a sense that all types of digital resources must be included. At Stage 1:

Organizational: There would be no explicit connection between institutional repositories and digital preservation programs and no explicit policies for either. Institutional repositories would be voluntary and opportunistic, taking advantage of any chance to acquire any digital content.

Technological: The technology infrastructure for each would be ad hoc, consisting of whatever is available or provided.

Resources: Resources for each would be finite and separate, not maximized.

Stage 2 Act. The motivation to move from Stage 1 to Stage 2 occurs when an organization perceives the need to take action to preserve its digital assets. Stage 2 activities are project-based, and often funded by external or one-time moneys, and the work tends to be conducted outside mainstream library functions. Although specifically addressing long-term issues, Stage 2 efforts tend to be of limited duration. IRs at this stage may focus on fulfilling the ingest and access goals that motivated their creation, without devoting much attention to preservation. This phase is usually the shortest in duration. At Stage 2:

Organizational: There would be separate implicit policies for institutional repositories and digital preservation representing some form of commitment.

Technological: The technology infrastructure for each would be project-based and therefore, hard to manage at the end of projects.

Resources: Resources for each would be minimal and project-based.

Stage 3 Consolidate. After some experience with parallel or sequential digital preservation projects, the organization generally concludes that the innate project lifecycle is not compatible with long-term planning and does not lead to the establishment of a program. Management of digital resources becomes ongoing and increasingly coordinated, but not yet truly integrated. At Stage 3 organizations realize that project-based funding is inadequate and unstable, and that a reliable, sustainable source of funding is needed to maximize the benefits of the work. Stage 3 is also characterized by the realization that something can be done now even as we wait for the big picture to emerge in full detail. A program mentality understands that investing in well-formed digital objects at creation assures that downstream those objects are easier to repurpose. At Stage 3:

Organizational: There would be separate policies for institutional repositories and digital preservation, but each would be explicit.

Technological: The technology infrastructure for each would be managed and moving toward coordination and joint investment.

Resources: Resources for each would be minimal, but ongoing.

Stage 4 Institutionalize. Bringing all of the pieces together across the institution allows for the best use of inevitably scarce human, technical, and financial resources and is the final internal step for the organization. Institutionalizing policies, procedures, and techniques creates a robust program that can be rationally managed and scaled as needs demand.

The motivation for moving from Stage 3 to 4 is the desire to maximize the effectiveness of resources through organization-wide efforts. The shift may be driven by the need to realize economies of scale through central or common, as opposed to individual, digital depository implementations. Organizations may linger at Stage 3 until a critical mass builds and the organization feels the need to move to the next stage. A driver for moving to Stage 4 may be the increasingly heavy burden of managing large, heterogeneous collections hosted separately.

Stage 4 programs exhibit organization-wide entities that coordinate, authorize, and mandate digital preservation mechanisms that allow for consistent and systematic management rather than event-based responses. The organization explicitly defines roles and responsibilities for key stakeholders. True technology planning and management begins, which is characterized by responding to rather than reacting to *and* anticipating needs. Investments in infrastructure are more likely to be based on requirements that are defined and approved at a high level of management and implemented across the organization.

Finally, rather than presuming that all digital materials will be preserved as part of the organization's commitment to digital preservation, the implications of that commitment are more fully understood and acceptance criteria are established and utilized to determine the scope of collections that will be actively preserved by the organization. Services to capture, store, maintain, and provide access to digital resources become integral to the organization and subject to relevant monitoring and measurements, and expectations that these services will be reliable and consistent become evident. At Stage 4:

Organizational: There would be one umbrella policy for digital content.

Technological: The technology infrastructure would be coordinated, jointly managed, responsive, anticipating technology developments to the extent possible.

Resources: Resources would be sufficient for managing institutional repositories and digital preservation programs.

Stage 5 Externalize. Stage 5 is characterized by inter-institutional collaboration. It may take the form of a consortium to build a digital archive, a federation of individual digital archives, or a virtual organization that comes together to manage one or more digital archives. Economies of scale, shared responsibility for infrastructure upkeep and pooled expertise are all possibilities. At this stage, the organization moves from discrete safe places as established at Stage 3, and integrated at the organizational level at Stage 4, to integrated safe places that bring multiple organizations, partners, and digital archive implementations together. Participation in subject-based, thematic, or domain-oriented depositories that cut across institutional lines may provide an impetus for moving from Stage 4 to 5.

Particularly for IRs, such a union of resources creates the potential and opportunity for a layer of services on top of the repositories that will be available to all of the members who may realize significant and unexpected benefits. In the ideal, this kind of success would both ensure the retention of existing partners and attract additional participants. The theme at this stage is that the whole *can* be greater than the sum of its parts. At Stage 5:

Organizational: The policies and management of institutional repositories and digital preservation programs would be collaborative.

Technological: The technological infrastructure would be distributed and coordinated.

Resources: Resources would be cumulative, inclusive of institutional repositories and digital preservation commitments and reflecting the designated resources of partners.

Envisioning the complete trajectory of organizational integration of institutional repositories and digital repositories is exciting, because suddenly next steps and goals come into clear focus. Understanding that digital preservation and institutional repositories will be segregated during Stage 1 (Acknowledge) and perhaps in Stage 2 (Act), managers can guide their institutional repository into a more modular relationship with digital preservation. The institutional repository may act as the Producer (pre-igest) and Consumer (access delivery) ends of the OAI model during Stage 2 (Act) and Stage 3 (Consolidate). Fully integrated institutional repositories and digital preservation programs at Stage 4 (Institutionalize) and Stage 5 (Externalize) exhibit a fully mature relationship, where an institutional repository's digital assets are managed and preserved as part of the organization's core functions.

CONCLUSION

Our paper concludes with suggestions for desirable and undesirable outcomes based on the five stages. These are some recommendations to leverage short-term benefits of institutional repositories to achieve long-term outcomes, rationalize efforts, and maximize resources:

- *Join forces:* Organizations should forge connections between institutional repositories and digital preservation programs to maximize resources, impact, and sustainability of each. Investing in both without coordinating these efforts will result in wasted resources—time, skills, and very often equipment. Organizations should consider the implications of having one without the other, especially an IR without a digital preservation program. Neither deposit nor storage equal preservation, and IRs risk embarrassment and disappointing faculty depositors if preservation services are not adequate.

- *Make commitments explicit:* If an institutional repository is intending to preserve the digital content, that intention should be expressed in the form of an explicit digital preservation policy. This commitment should be accessible to and understandable by depositors. This is so because the existence of an IR does not necessarily entail a digital preservation commitment—some do intend to preserve and some do not. An IR may be only one entry point for bringing digital content into a digital preservation program and not all digital content deposited in an IR may need to be preserved. If an IR does not have or intend a preservation mandate, that should be made explicit to depositors.
- *Manage expectations:* Not all digital content accepted by institutional repositories may warrant long-term preservation. Establishing clear, known selection criteria for each will manage expectations and avoid implicit and unfunded mandates. IRs and digital preservation programs should coordinate selection criteria, adjust as needed, and handle exceptions.
- *Provide tools and guidelines:* Providing tools and guidance for creating well-formed, archive-ready digital content will benefit all involved—producers, consumers, archivists, and institutional repository managers. All opportunities to raise awareness and enable good digital asset management should be pursued.
- *Provide seamless interfaces and services:* An integrated organizational approach would provide an approachable and extensible front end for producers and consumers using institutional repositories as a service layer, with a reliable and proven sustainability approach for digital content. Users should not have to know who is doing what behind the interface.
- *Embrace OAIS as a collaborative model:* Mapping the components of the institutional repository and the digital preservation program will make overlaps and gaps clear. OAIS defines full lifecycle management for digital content of all kinds. The roles, functions, and content definitions of OAIS provide effective management tools for integrating IRs and digital preservation programs.
- *Emphasize strengths of each:* The skills required for implementing institutional repositories enable interactions with producers. The skills required for digital preservation programs enable the appraisal of digital content, the mapping of digital content requirements to appropriate digital preservation strategies, the identification of potential preservation challenges presented by content ingested into IRs, and an awareness of digital preservation standards and practices.
- *Incorporate institutional records:* Institutional records are a specialized form of digital content that could and often should be incorporated into managed digital content programs. Institutional records are the responsibility of the archival community, yet electronic records programs

have been slow to flourish in academic environments. Beyond published sources, bringing institutional records in through institutional repositories may expand the reach of archival programs and of IR content making both more sustainable. Archivists and librarians joining forces to ensure the preservation of institutional records could be a strong partnership—the skills and organizational networks of each should be applied to the problem.

- *Reduce then remove barriers between digital preservation and institutional repository programs as programs mature.* Eventually, there should be no need for distinguishing the institutional repository from the digital preservation program. Integrating the two should be possible as the collaboration matures. When Stage 5 is achieved, an organization's IR and digital preservation program will each be an integral part of a well-managed whole. Integration should not be rushed or both may experience negative impacts, for example, preservation requirements perceived as disincentives to depositors and unfiltered ingest for IRs as an unmanageable preservation burden.
- *Focus on well-managed collections to succeed at both:* Focusing on the requirements and well-being of the digital content rather than on organizational barriers and challenges will make it easier to meld the objectives of institutional repositories and digital preservation programs.

As we focus on pulling our digital treasure through the dangers wrought by the passage of time—changing organizations, technologies, and budgets—there are enormous benefits to be gained by yoking digital preservation programs with institutional repositories. With good management and cooperation, we stand a good chance of achieving systems where information is both accessible and safely kept.

ACKNOWLEDGMENT

The authors wish to thank Anne R. Kenney who was a codeveloper of the Five Stages of digital preservation programs analysis. She provided valuable comments and her insights contributed significantly to the paper. Any errors are, of course, our own.

REFERENCES

- Archival Workshop Program Committee. (1999). *Draft report of the Archival Workshop on Ingest, Identification, and Certification Standards (AWICS)*. College Park, MD, October 13–19, 1999. Retrieved June 19, 2008, from <http://nost.gsfc.nasa.gov/isoas/awiics/>
- Budapest Open Access Initiative. (2002). Retrieved June 13, 2008, from <http://www.soros.org/openaccess/read.shtml>
- Consultative Committee for Space Data Systems (CCSDS). (2002). *Open Archival Information System, (OAIS), ISO 14721:2003, CCSDS Blue Book*. Washington, DC: CCSDS Secretariat. Retrieved June 13, 2008, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Davis, P. M., & Connolly, M. J. L. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13(3/4). Retrieved June 25, 2008, from <http://www.dlib.org/dlib/march07/davis/03davis.html>

- Deutsche Initiative für Netzwerkinformation (*DINI*) Working Group on Electronic Publishing. (2003, November). *DINI*-Certificate Document and Publication Repositories, *DINI*-Schriften 3-en. Now superseded by Version 2.0 September 2006. Retrieved June 19, 2008, from <http://edoc.hu-berlin.de/series/dini-schriften/2006-3-en/PDF/3-en.pdf>
- Digital Curation Centre (DCC) & DigitalPreservationEurope (DPE). (2007). *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) Version 1.0 (draft)*. Retrieved June 13, 2008, from <http://www.repositoryaudit.eu/>
- DCC, DPE, nestor, & CRL. (2007). Core requirements for digital archives. Retrieved June 19, 2008, from <http://www.crl.edu/content.asp?11=13&12=58&13=162&14=92>
- Digital Repository Audit & Certification Birds of a Feather Group. (2008). Digital repository audit & certification wiki. Retrieved June 13, 2008, from <http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/WebHome>
- DSpace. (2008). DspaceInstances. *DSpace wiki*. Retrieved June 13, 2008, from <http://wiki.dspace.org/index.php/DSpaceInstances>
- Estlund, K., & Neatrou, A. (2007). Utah digital repository initiative. *D-Lib Magazine*, 13(11/12). Retrieved June 13, 2008, from <http://www.dlib.org/dlib/november07/neatrou/11neatrou.html>
- Ferreira, M., Rodriguez, E., Baptista, A. A., & Saraiva, R.. (2008). Carrots and sticks: Some ideas on how to create a successful institutional repository. *D-Lib Magazine*, 14(1/2). Retrieved June 13, 2008 from, <http://www.dlib.org/dlib/january08/ferreira/01ferreira.html>
- Guterman, L. (2008, May 7). Harvard Law School votes for open access to scholarship. News Blog: *Chronicle of Higher Education*. Retrieved June 24, 2008, from <http://chronicle.com/news/article/?id=4445>
- Harnad, S. (1999). Free at last: The future of peer-reviewed journals. *D-Lib Magazine*, 5(12) Retrieved June 13, 2008, from <http://www.dlib.org/dlib/december99/12harnad.html>
- Hitchcock, S., Brody, T., Hey, J. M. N., & Carr, L. (2007). Digital preservation service provider models for institutional repositories: Towards distributed services. *D-Lib Magazine*, 13(5/6). Retrieved June 13, 2008, from <http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html>
- ICPSR. (2007). Digital preservation management. Retrieved November 11, 2008, from <http://www.icpsr.umich.edu/dpm/>
- Kenney, A. R., & Buckley, E. L. (2005). Developing digital preservation programs: The Cornell survey of institutional readiness, 2003–2005. *RLG DigiNews*, 9(4). Retrieved September 25, 2008, from <http://worldcat.org/arviewer/1/OCC/2007/08/08/0000070511/viewer/file3612.html>
- Kenney, A.R., Entlich, R., Hirtle, P. B., McGovern, N. Y., & Buckley, E. L. (2006) *E-Journal archiving metes and bounds: A survey of the landscape*. Washington, DC: Council on Library and Information Resources. Retrieved June 25, 2008, from <http://www.clir.org/pubs/abstract/pub138abst.html>
- Kenney, A. R., & McGovern, N. Y. (2003). The five organizational stages of digital preservation. In P. Hodges, M. Bonn, M. Sandler, and J. P. Wilkin (Eds.), *Digital libraries: A vision for the 21st century: A festschrift in honor of Wendy Lougee on the occasion of her departure from the University of Michigan*. Ann Arbor, MI: University of Michigan Library, Scholarly Publishing Office.
- Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, 226. Retrieved June 13, 2008, from <http://dspace.uniroma2.it/dspace/bitstream/2108/261/1/ir.html>
- McDowell, C. S. (2007). Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, part 2. *D-Lib Magazine*, 13(9/10). Retrieved June 13, 2008, from <http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html>
- National Institutes of Health. (2008). NIH public access policy. Retrieved June 24, 2008, from <http://publicaccess.nih.gov/policy.htm>
- National Science Foundation Cyberinfrastructure Council. (2007). *Cyberinfrastructure vision for 21st century discovery*. National Science Foundation. Retrieved June 24, 2008, from <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>
- nestor. (2006). Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version 1 (Entwurf zur öffentlichen Kommentierung). Edited by nestor—Network of Expertise in Long-Term Storage of Digital Resources and nestor Working Group on Trusted Reposi-

- ories Certification (Frankfurt am Main: nestor). Retrieved November 10, 2008, from <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>. English translation: nestor. (2006). Criteria for Trusted Digital Long-Term Preservation Repositories - Version 1 (Request for Public Comment). Edited by nestor—Network of Expertise in Long-Term Storage of Digital Resources and nestor Working Group on Trusted Repositories Certification (Frankfurt am Main: nestor) nestor materials 8. Retrieved November 10, 2008, from <http://nbn-resolving.de/urn:nbn:de:0008-2006060703>
- Okerson, A. S., & O'Donnell, J. J. (Eds.). (1995). *Scholarly journals at the crossroads: A subversive proposal for electronic publishing: An Internet discussion about scientific and scholarly journals and their future*. Washington, DC: Association of Research Libraries, Office of Scientific and Academic Publishing. Retrieved June 6, 2008, from <http://www.arl.org/bm-doc/subversive.pdf>
- PREMIS. (2007). PREMIS implementation registry. Retrieved June 5, 2008, from <http://www.loc.gov/standards/premis/premis-registry.php>
- PREMIS Editorial Committee. (2008). *PREMIS Data Dictionary for Preservation Metadata, version 2.0*. Retrieved June 13, 2008, from <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>
- Prom, C., & E. Swain. (2007). From the college democrats to the falling Illini: Identifying, appraising, and capturing student organization websites. *American Archivist*, 70(2), 344–363.
- Rieh, S. Y., Markey, K., St. Jean, B., Yakel, E., & Kim, J. (2007). Census of institutional repositories in the U.S. *D-Lib Magazine*, 13(11/12). Retrieved June 13, 2008, from <http://www.dlib.org/dlib/november07/rieh/11rieh.html>
- Rieger, M. (1966). Archives and automation. *American Archivist*, 29(1), 109–111.
- RLG/OCLC. (2002). *Trusted digital repositories: Attributes and responsibilities*. Mountain View, CA: Research Libraries Group. Retrieved June 13, 2008, from <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>
- RLG-NARA Digital Repository and Certification Task Force. (2007). *Trustworthy repositories audit & certification: Criteria and checklist version 1.0 February, 2007*. OCLC & CRL. Retrieved June 19, 2008, from <http://www.crl.edu/PDF/trac.pdf>
- SCOAP³—Sponsoring Consortium for Open Access Publishing in Particle Physics. (2008, June 23). Three more U.S. labs and two more U.S. universities join SCOAP³. Retrieved June 24, 2008, from <http://www.scoap3.org/news/news43.html>
- SHERPA Core Team. (2008). *SHERPA-RoMEO*. Nottingham, UK: University of Nottingham. Retrieved June 13, 2008, from <http://www.sherpa.ac.uk/romeo.php>
- Thomas, C., & McDonald, R. H. (2007). Measuring and comparing participation patterns in digital repositories: Repositories by the numbers, part 1. *D-Lib Magazine*, 13(9/10). Retrieved June 13, 2008 from <http://www.dlib.org/dlib/september07/mcdonald/09mcdonald.html>
- Treloar, A., Groenewegen, D., & Harboe-Ree, C. (2007). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine*, 13(9/10). Retrieved June 13, 2008, from <http://www.dlib.org/dlib/september07/treloar/09treloar.html>
- Waters, D., & J. Garrett. (1996). Preserving digital information: Report of the task force on archiving of digital information. *CLIR Report*, 63. Retrieved September 6, 2008, from <http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>

In September 2006, Nancy Y. McGovern became the digital preservation officer at the Inter-university Consortium for Political and Social Research (ICPSR). For the five years prior to that, she was the director of Research and Assessment Services and digital preservation officer at Cornell University Library. She has focused on digital preservation research and practice since 1986, when she began a decade of service on the senior staff of the Center for Electronic Records at the U.S. National Archives. She cofounded the Society of American Archivists' Research Forum in 2007. She is completing her PhD on a digital preservation topic through University College London. She served as coeditor of *RLG DigiNews* from 2001 to 2006 and co-developed the Digital Preservation Management workshop series and tutorial with Anne R. Kenney beginning in 2002.

Aprille Cooke McKay is a former Digital Preservation Specialist at ICPSR, where she managed the Digital Preservation Workshop program and served as a workshop instructor. Previously, she was the project manager for the Mellon-funded project, "Developing Standardized Metrics in College and University Archives and Special Collections," based at the University of Michigan's School of Information, working with Ax-SNET investigators Elizabeth Yakel, Helen Tibbo, and Wendy Duff. She holds a BA from the University of Virginia, a JD from the University of Chicago, and an MSI with specialization in Archives and Records Management from the UM School of Information. She has worked at the Bentley Historical Library, specializing in legal collections, and has also practiced law in the Chicago area. She is a member of SAA's Intellectual Property Working Group, Website Working Group, and the Standards Committee.