Aligning National Approaches to Digital Preservation

Nancy Y. McGovern, Volume Editor
Katherine Skinner, Series Editor

Aligning National Approaches to Digital Preservation

Nancy Y. McGovern, Volume Editor

Katherine Skinner, Series Editor

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ECONOMIC ALIGNMENT

Maurizio Lunghi (Fondazione Rinascimento Digitale)
Neil Grindley (Joint Information Systems Committee)
Bohdana Stoklasová (National Library of Czech Republic)
Aaron Trehub (Auburn University, ADPNet)
Christin Egger (Fondazione Rinascimento Digitale)

Abstract

This essay presents an overview of the economic issues that define, promote, or inhibit effective national and international programs for preserving digital cultural heritage materials. Specifically, it presents and discusses multi-institutional approaches to identifying effective and economically sustainable policies in managed digital information environments, citing current digital preservation initiatives in North America and Europe. The essay will also address related issues, including: service/user relationships, roles and responsibilities throughout the various communities, the choice of suitable business models, and cost analyses as essential components of defining economically sustainable approaches to preservation. In keeping with the aims of the Aligning National Approaches to Digital Preservation conference, the essay concludes by considering what a blueprint for success in this area might look like and offering specific recommendations to that end.

Introduction

Economic sustainability—that is, generating and allocating the resources necessary to support long-term preservation activities—is fundamental for the success of long-term digital preservation programs. If there is disagreement about other aspects of digital preservation, there should at least be agreement on this. And yet, this fundamental point has not received the attention or the analysis it deserves. As the authors of the final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010, hereafter referred to as BRTF-SDPA Final Report) have pointed out, while there is a substantial literature on the technical and policy aspects of digital preservation, the economic aspects have, until quite recently, been relatively neglected.

The authors of this essay hope to help remedy this gap by focusing on questions of economic alignment and economic sustainability as they affect digital preservation initiatives in North America and Europe. The essay reflects the views and experiences of its authors, but it also incorporates the results of discussions at the 2010 "Aligning National Approaches to Digital Preservation" (ANADP) conference in Tallinn, Estonia. The conference organizers posed three general questions to the participants, with a view to formulating an action plan for the international digital preservation community. The first question was to consider the most important alignment accomplishments that have taken place in the digital preservation field. The second was to examine the current challenges and gaps that represent barriers to establishing sustainable digital preservation activities. And the third asked where the panel thought the digital preservation community should plan to be in five years' time and what would success in this area look like? These questions conveniently encompass many of the issues that have a bearing on economically sustainable digital preservation strategy and action. There are a number of additional questions and issues, however, that relate more specifically to economic alignment. They include: the nature of costs and business models, the effectiveness and demand for services, strategies for selection and appraisal, requirements for partnership and training, and the general need for clarity around roles and responsibilities.

Digital Preservation: A Value Proposition

The long-term preservation of digital materials is an issue that has global relevance. It has become generally understood since the publication in 1996 of the landmark Garrett-Waters report on preserving digital information that engagement with preservation is an unavoidable corollary to the creation and use of nearly all forms of digital content.

Individuals, organizations and businesses are usually highly motivated to think about the issues and challenges they are likely to face in the next phase of planning, which generally means three to five years. Five years is not that long in digital preservation terms, however, and this short-term perspective is probably the single most critical reason that making a business case or economic argument for preservation is a difficult proposition. Therefore, the first and most important concept to argue is that some digital information *does have implicit enduring value*; or alternatively, that it can be used to create entities that will have value. Whilst a

case can and should more often be made for the short-term benefits of preservation, it is this long-term value proposition that underpins all other arguments and evidence for engagement in this area. Digital preservation often looks to its equivalent in the physical realm and cites the maintenance of manuscripts over centuries as proof of the impact and worth of caring about the integrity, complexity, intricacy and context of materials produced by human endeavor. The starting point for this paper, therefore, is that digital preservation is an important activity that will enable this generation and subsequent generations to make choices and exploit opportunities that they would otherwise be unable to take advantage of. It is ultimately these human outcomes, rather than technical or bureaucratic ones, that make the economic and every other case for digital preservation.

Putting the ideological view to one side and given the understandable focus of most people on short-term goals, persuading a wide range of information professionals that digital preservation ought to be an essential and embedded part of their daily work is always going to be a challenge. Given that issues span technical, legal, educational, organizational, and of course economic categories, there is an innate complexity to tackling digital preservation that many find a disincentive to engagement. For the minority that find this complexity stimulating, digital preservation continues to present rewarding intellectual opportunities. For the vast majority, however, continuing "access to" or future "use of" the preserved materials will always be the principle motivation for continuing to fund preservation activity. This level of interest from the user community is crucial. Preservation, whether physical or digital, is going to seem like wasted investment without any current or future usage intention. If the demand for access to preserved digital objects and their permanent storage is well articulated, then economic sustainability becomes far more likely. If those arguments originate from the community, and even across national boundaries, then so much the better.

The difficulty of assigning accurate value to digital information is a global problem and sharing that problem is a good mitigation measure. Whilst it may be possible in hindsight to judge that people made errors of judgment in assigning substantial resources to preserving material that was subsequently never used or was considered of negligible value, it will be a compelling

defense to cite community, national, or international precedent as proof of good faith.

Economic Alignment: Core Approaches

The first core task is to consider where progress has already been made, either nationally or internationally, to help ameliorate problems relating to the economics of digital preservation. The topic itself encapsulates a lot of complexity in that there are various perspectives that need to be factored into any discussion of what constitutes economic issues in this field. The focus could conceivably be on the cost of maintaining digital material over time, the budget strategies of organizations obliged to engage with preservation, the economic framework in which preservation may effectively occur, or the type and extent of funding required for effective preservation to flourish. The aforementioned BRTF-SDPA Final Report represents a notable accomplishment in this area, in the sense that it was the first—and to date the only—comprehensive examination of digital preservation from an economic perspective. Among other things, the report offered a succinct definition of economically sustainable preservation[1] and analyzed the economic factors involved in the preservation of four types of digital content: scholarly discourse, research data, commercially owned cultural content, and collectively produced Web content. (We would add a fifth category to this list: digital content produced and owned by libraries, archives, museums, and other cultural heritage organizations.) In the course of disentangling and classifying the different elements of digital preservation, the report's authors posited five conditions for sustainable digital preservation:

1. Recognition of the benefits of preservation by decision makers;

2. A process for selecting digital materials with long-term value;

3. Incentives for decision makers to preserve in the public interest;

4. Appropriate organization and governance of digital preservation activities; and

---

[1] "[A] means of keeping information accessible and usable over time by ensuring the ongoing and efficient allocation of resources to its maintenance" (BRTF-SDPA Final Report, 2010, p. 107).

5. Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities (p. 12).

This paper takes the view that all of these perspectives and the criteria for economic sustainability are valid areas for discussion, though some have been subject to more development and attention than others in terms of the amount of alignment that may have occurred.

The first of those options—the work that has been done on the lifecycle cost of information management—is arguably the most widely understood interpretation of any question about the "economics of preservation" and probably makes the most immediate sense to the non-specialist who may be concerned to know whether preservation constitutes a "nice-to-have" but dispensable layer of assurance, or whether it is an information management necessity. Knowing the cost of preservation does not necessarily decide this question, but it may focus the enquirer's mind on how seriously he or she needs to contemplate the question.

In the United Kingdom, the cost of preservation has recently been the focus of various phases of the LIFE project[2] undertaken by the British Library and University College London. The project developed and refined a lifecycle model that primarily relates to materials that may be found in a (digital) library context (e.g. text and images) and developed a complex spreadsheet tool to help with calculating the cost over time of storing, managing, and preserving that material. This work has also been picked up and further developed by the Danish National Archives and the Royal Library of Denmark[3] and an online version of the costing tool is being developed and piloted by the Humanities Advanced Technology & Information Institute (HATII) at the University of Glasgow in collaboration with the Open Planets Foundation (OPF). Further detailed work looking at the long-term cost of preserving materials, in this instance research data, was carried out in the first two phases of reporting by the Keeping Research Data Safe (KRDS) project (the third and final phase defined a taxonomy for identifying direct and indirect benefits of long-term digital

---

[2] Lifecycle Information for E-Literature: http://www.life.ac.uk/ (last accessed 03-08-2012).

[3] Anders Bo Nielsen and Ulla Bøgvad Kejser. 2008. "Using the LIFE Costing Model: Case Studies from DK." Available at: http://discovery.ucl.ac.uk/9313/ (last accessed 03-08-2012).

preservation).[4] As well as relying on new research in collaboration with data centers to assess the real costs of keeping data over long periods, the KRDS project drew on both the LIFE Project modeling work and the Cost Estimation Tool (CET) developed by NASA, and other resources such as: the TRAC (Transparent Approach to Costing) Model, the Open Archival Information System (OAIS) Reference Model, and the Digital Curation Centre (DCC) Lifecycle Model in order to create an effective generic framework to discover the cost of managing research data. More generally, the cost of digital preservation figured prominently in the eSpida Project at the University of Glasgow, an initiative aimed at "exploring how intangible assets might be valued in order to make a sound business case to ensure the longevity of information objects; in other words, achieve truly sustainable digital preservation" (Currall and McKinney, 2006). In the United States, Serge Goldstein and Mark Ratliff (2010) of the Office of Information Technology at Princeton University have devised a cost model for the long-term preservation of research data. Dubbed "DataSpace," the model includes a "Pay Once, Store Forever" (POSF) funding formula. The price schedules for community-based digital preservation initiatives like HathiTrust, Chronopolis, and the Private LOCKSS Networks as well as proprietary preservation services like Portico and Tessella also embody assumptions about the cost of long-term preservation.

So the stewardship cost of keeping digital material over time has been demonstrably addressed by various projects, both recently and in the past, and it seems appropriate to declare that some alignment around this work, and the initiatives of other organizations and projects on this topic, has taken place—if not around the precise cost of various preservation tasks, then at least around some of the digital lifecycle information models on which they are based. These models are themselves significant as the digital equivalent to earlier examples from the realm of archival practice and records management, the former a discipline that goes back hundreds of years and the latter an activity that emerged in response to the burgeoning amount of documentation being produced during the middle of the 20th century. There is plenty of evidence to suggest that the lifecycle of information and its management is well understood by now, and there is also reason to

---

[4] Keeping Research Data Safe (KRDS) Web site: http://www.beagrie.com/krds.php (last accessed 03-08-2012).

believe that the main components of digital preservation have been successfully described and categorized. Whether every component in a diagram such as the DCC Lifecycle model[5] is understood and implementable (or even practical to contemplate) for many organizations is another question, but there does appear to be some alignment and agreement about the nature of, and the relationship between, preservation tasks.

Slightly more contentious, particularly beyond the edges of the broad preservation community, is the notion that there is alignment or consensus around the principle of appraisal and selection. This is a deeply significant point in relation to the economics of preservation since the amount of material that one chooses to keep does, of course, have an impact on the infrastructure that one needs to manage it—a point made repeatedly in the BRTF-SDPA Final Report. It is indeed true that at least amongst communities that have spent time thinking hard about the consequences of information management policies (economic and otherwise), there is alignment about the value— indeed the necessity—of selecting and appraising digital information: in effect, assigning value to it and prioritizing some data as more valuable than others. There is, however, less alignment about the practicality and processes for actually carrying out selection and appraisal routines. This point will be addressed in the "Gaps and Challenges" section below.

Another highly visible area of alignment that must surely result in enhanced economic sustainability for digital preservation is the amount of community-building and the national and international collaborations that have occurred, not only as a result of the numerous seminars, workshops and conferences that take place around the world, but also from the open exchange of information that takes place between preservation practitioners, many of whom are based within public and non-profit institutions such as universities, libraries, and archives. Whilst it would be banal to spell out the benefits of cooperation and discussion between theorists and practitioners in any given field, the exchange of experience and good and bad practice; wide participation in advocacy and awareness raising; and the development of common terminology and common approaches have all been key components of establishing digital preservation as a sub-discipline.

---

[5] DCC Lifecycle model, see http://www.dcc.ac.uk/resources/curation-lifecycle-model (last accessed 03-08-2012).

International cooperation has not simply been a by-product or an extension of the peer-review process: it has been critical for the establishment of practice and policy in a field where many onlookers are still waiting to hear and understand what a convincing and robust long-term business case for preservation looks and sounds like.

As well as providing opportunities for forming useful contacts and becoming more closely acquainted with the concerns of peer practitioners, attending and participating in meetings is a way of accelerating the learning and training process for staff who are developing knowledge in the field. This is of very practical economic benefit to organizations that might otherwise have to contemplate expensive training and staff development. National and international preservation-related conferences, workshops, seminars, symposia, and other events are numerous, occasionally free, and increasingly focused on communicating and delivering practical preservation outcomes.

In addition to standalone or annual events such as International Conference on the Preservation of Digital Objects (iPres), and International Digital Curation Conference (IDCC), funded projects have made an enormous contribution to aligning policy, strategy and practice in the field, not only through dissemination meetings funded as part of project work plans, but also through their associated reports and deliverables. One of the outstanding contributions in this area has been made by the European Commission, which has funded a number of major European projects that continue collectively to have a significant impact on digital preservation. These include the following:

- ERPANET: Electronic Resource Preservation and Access Network

- DPE: Digital Preservation Europe

- PLANETS: Preservation and Long-Term Access through Networked Services

- CASPAR: Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval

- KEEP: Keeping Emulation Environments Portable

- PrestoSpace: Preservation Towards Storage and Access. Standardised Practices for Audiovisual Contents in Europe.

- PARSE Insight: Permanent Access to the Records of Science in Europe

- APARSEN: Alliance for Permanent Access to the Records of Science Network

These are all ambitious multi-partner institutional undertakings where many participants from all over Europe (and in some cases beyond) have been given an opportunity to hone or develop their skills in an emerging area. Whilst it is not training as such, there will almost certainly have been ample requirement for many participants to learn quickly "on the job," and this accelerant factor, bringing people up to speed within finite deadlines, is of broad economic benefit.

Similar work is being carried out at the national level in the United States under the auspices of the National Digital Information Infrastructure and Preservation Program (NDIIPP) of the Library of Congress. The NDIIPP's mission is "to develop a national strategy to collect, preserve and make available significant digital content, especially information that is created in digital form only, for current and future generations,"[6] and to that end it has focused on three areas: capturing, preserving, and making available digital content; building a nationwide network of preservation partners, including the MetaArchive Cooperative and the Chronopolis digital preservation network; and directly supporting or promoting a technical infrastructure of tools and services, including BagIt, Heritrix, and the JSTOR/Harvard Object Validation Environment (JHOVE). Perhaps the NDIIPP's most important accomplishment to date has been articulating a convincing case for the importance of long-term digital preservation, one that bears the imprimatur of the closest thing that the US has to a national library. An endorsement by the Library of Congress carries weight for organizations working in related fields and the Library has succeeded at least in making the argument that digital preservation ought to be a national priority. This can be seen, for example, on the Web site for the National Digital Stewardship Alliance (NDSA[7]), an outgrowth and extension of the NDIIPP.

---

[6] NDIIPP Web site: http://www.digitalpreservation.gov/ (last accessed 03-08-2012).
[7] NDSA Web site: http://www.digitalpreservation.gov/ndsa (last accessed 03-08-2012).

In addition to the Library of Congress, the Institute of Museum and Library Services (IMLS), a federal funding agency, has also supported digital preservation initiatives in the United States, most notably the establishment of the Alabama Digital Preservation Network (ADPNet), a statewide LOCKSS-based network. Aaron Trehub discusses this initiative and the challenge of achieving economic sustainability elsewhere in this volume.

Finally, the National Science Foundation (NSF), the National Archives and Records Administration (NARA), and the National Historical Publications and Records Commission (NHPRC: a unit of NARA) have supported work on "implementing solutions to the challenges of preserving electronic records with permanent historical value."[8] This work includes the development of the Integrated Rule Oriented Data System (iRODS) at the Data Intensive Cyber Environments (DICE) Center at the University of North Carolina at Chapel Hill and the Institute for Neural Computation at the University of California San Diego.

Despite these tangible proofs of progress, however, it can be incredibly difficult to persuade library administrators and other decision makers to embrace the requirements of digital preservation and to get it embedded into organizational strategies and thought processes. Bohdana Stoklasová addresses some of the challenges of advocating for preservation at these levels later in this volume. She argues that the gradual introduction of both effective technology and skilled personnel is a critical requirement, but it is not cheap or easy to accomplish.

Once momentum is achieved, however, and the backing of powerful advocates secured, a great deal of progress can be made and partnerships can be brokered and usefully exploited. Returning again to North America, the Library of Congress, IMLS, and other funding organizations have supported efforts to define best practices and procedures for digital preservation. They have also supported the development of governance instruments (a crucial but often-overlooked precondition for creating economically sustainable and scalable preservation networks, especially among different kinds of institutions in different states, provinces, and countries), and have actually created functioning preservation networks. For example, the NDIIPP supported the creation of the

---

[8] See http://www.archives.gov/nhprc/apply/eligibility.html (last accessed 03-08-2012).

Data-PASS network; the Persistent Digital Archives and Library System (PeDALS) project; and the MetaArchive Cooperative, which was the first Private LOCKSS Network (PLN) explicitly designed for the preservation of locally created (and locally owned) digital content. For its part, the IMLS-supported ADPNet was the first statewide PLN and served as the model for the Council of Prairie and Pacific University Libraries (COPPUL) PLN in western Canada. Indeed, the ADPNet-COPPUL relationship represents a working example of economic alignment and offers proof that it is possible to create affordable and sustainable preservation networks in very different environments.

In the UK, the Joint Information Systems Committee (JISC) has been influential in funding innovation and building capability through preservation programs and projects (most often based within UK universities) that have supported a wide range of activity including feasibility and scoping work, technical development, policy and legal studies, and network and partnership support. The Dutch National Library and the National Archives have been an influential force in the Netherlands driving preservation practice there and being influential around the world, as have their UK, Australian, New Zealand, German, and Danish counterparts (in association with those responsible for their core and capital funding). It is worth noting that this partial and arbitrary list exclusively describes publicly funded organizations and this goes some way to underpin the next point of alignment, which is around the theme of "openness." It is tempting to think that the natural tendency of all publicly funded organizations would surely be towards the open: i.e. open source (software), open access (content), open standards, and indeed open communities, where participants from all sectors are welcome and encouraged to join in the discussion and add value. But on reflection, this is taking an influential core value of a group of mostly large and powerful institutions and extending it across a whole diverse community.

Intuitively, the adoption of "open" approaches, in particular open-source software in the context of technical preservation solutions, seems like a tactic designed to appeal to cash-strapped organizations with little by way of resources and funding to engage with the complexities of preservation. But as anyone who works with a range of software will state, "open source does not mean free." Whilst the source code may be accessible and reusable, there may be a cost for distribution, for support, or for particular license

conditions. Additionally, to actually implement, use, and locally maintain the software in a way that is useful for one's own organization, there may well be significant costs down the line that are inherent to a community-owned resource. In some cases, it may be valid to argue that such costs would be usefully subsumed within a service-level agreement on a piece of proprietary software from a commercial vendor. That said, there has nonetheless been great progress in establishing an array of open and free tools, toolkits, models, frameworks, and guidance that have removed many of the financial barriers to engaging with preservation, at least up to a certain level. Resources such as Archivematica (a comprehensive digital preservation system); the California Digital Library's Data Management Planning Tool (DMP Tool) and suite of microservices for data curation; The Curator's Workbench (University of North Carolina at Chapel Hill pre-ingest tool); and HOPPLA (Vienna University of Technology) may indicate the path to further progress in this direction. Other tools such as DRAMBORA (a risk audit tool), AIDA (Assessing Institutional Digital Assets), and DAF (the Data Asset Framework) are being combined in an integrated suite to tackle sophisticated work in the area of long-term data management practice.[9] This approach emulates various projects over the years that have built on and integrated various robust preservation components such as DROID and PRONOM (The UK National Archives); JHOVE (JSTOR and the Harvard University Library); and the NLNZ Metadata Extractor (National Library of New Zealand).

In terms of open standards, probably the most widely referenced and influential standard reference point in Digital Preservation is ISO 14721, better known as the Open Archival Information System (OAIS) reference model.[10] The OAIS model is an excellent framework for defining the inputs, processes, and outputs of an eligible preservation system and it is this flexibility, combined with a collection of canonical terms and an easily reproducible graphic, which has earned it a ubiquitous role throughout the preservation literature and a place in almost every entry-level presentation ever given about the topic. But alongside

---

[9] See the IDMP: Integrated Data Management Planning & Support Project at the DCC:
http://www.jisc.ac.uk/whatwedo/programmes/mrd/supportprojects/idmpsupport.aspx (last accessed 03-08-2012).

[10] ISO (International Organization for Standardization) 14721:2003; originally proposed by the Consultative Committee for Space Data Systems.

its usefulness, particularly in the areas of teaching and training, the OAIS model has an economic relevance precisely because it is an open and free framework. One of the great benefits of the model is that it describes a workflow and environment that adheres to good and effective working practices without being too prescriptive about compliance with detailed implementations of (and therefore investments in) particular forms of infrastructure.

Economic Alignment: Gaps and Challenges

The alignment accomplishments alluded to above signal that preservation, and more particularly economic positions in relation to preservation, have either purposefully been developed (e.g. cost models, dedicated preservation funding programs, and institutional policy development) or have realized some oblique benefits from the tendency towards "openness" in many public institutions, and also perhaps the general move towards e-only provision of resources (a trend that is particularly apparent in the area of scholarly journals).

But despite the many advances there is still a great deal to be done to ensure that we have sustainable economic strategies for preservation. This is especially important precisely because digital preservation can be a financially demanding undertaking whose benefits may not be immediately apparent. A large number of digitized volumes is eye-catching proof of a project's "success;" the substantial investment required to ensure their long-term availability is invisible to users and less likely to generate enthusiasm among decision makers. As a result, long-term preservation is still not perceived as an indispensable part of digitization projects and its cost is often underestimated or entirely ignored in favor of digitizing more materials. The ability to build effectively on previous and current investment is therefore a powerful argument for digital preservation and evidence of its economic sustainability.

Building on Current Investment

Given the wide variety of institutions that should be concerned about digital preservation and the differences among them in culture, policies, laws, regulations, and resource levels, it would be misleading to speak of economic alignment in terms of a single, uniform approach. What may work for one type of institution in a given country would not work at all for the same type of institution in another country. That said, there are general

principles that are useful in designing economically sustainable digital preservation networks, and some of them can be discerned from working examples in North America and Europe. One task, therefore, might be to compile an international library of recommended practices that can be modified and applied to different situations; in other words, national resource pages writ large.

Fortunately, there are a growing number of working examples to draw on, some of them of fairly lengthy provenance. The San Francisco-based Internet Archive (IA) was founded in 1996 as a non-profit organization by Internet entrepreneur Brewster Kahle and now contains almost five million texts, moving images, live music concerts, and audio recordings. A number of Private LOCKSS Networks (PLNs) have been established in North America, the United Kingdom, and Germany in order to preserve locally created digital content.[11] The oldest of them, the MetaArchive Cooperative, was founded in 2004 under the auspices of the NDIIPP and now numbers almost fifty member institutions in the United States, the United Kingdom, Brazil, and Spain. The aforementioned Integrated Rule Oriented Data System (iRODS) was launched in 2006; it is now based at the Data Intensive Cyber Environments (DICE) Center at the University of North Carolina at Chapel Hill and the Institute for Neural Computation at the University of California San Diego and is in use at the Carolina Digital Repository, the Texas Digital Library, the Bibliothèque nationale de France (BnF), and other cultural heritage organizations in the United States and Europe. One iRODS-based network, Chronopolis, is based at the San Diego Supercomputer Center and the University of California San Diego and offers a suite of distributed preservation services to other institutions. HathiTrust is a large-scale collaborative repository of digitized content from the Google Books initiative, the Internet Archive, and local digital collections. Established in 2008 by the thirteen member libraries of the Committee on Institutional Cooperation (CIC), the HathiTrust partnership now includes over sixty research libraries in the United States and Europe and has

---

[11] Examples include the MetaArchive Cooperative (http://www.metaarchive.org); the Alabama Digital Preservation Network (ADPNet: http://www.adpn.org/); the Council of Prairie and Pacific University Libraries Network (COPPUL: http://coppullockssgroup.pbworks.com/); the UK LOCKSS Alliance (http://www.lockssalliance.ac.uk/); and LuKII (http://www.lukii.hu-berlin.de/) (all last accessed 03-08-2012).

digitized almost ten million volumes, almost 30 percent of which are in the public domain. The University of California Curation Center (UC3) recently launched Merritt, a digital repository and preservation service for the University of California community. Finally, in November 2011, DuraSpace—a non-profit organization formed by the merger of DSpace and Fedora—announced DuraCloud, a cloud-based service aimed at "providing preservation support and access services for academic libraries, academic research centers, and other cultural heritage organizations." Among those organizations are Hamilton College, Indiana University-Purdue University Indianapolis, the Inter-University Consortium for Political and Social Research (ICPSR), and the Massachusetts Institute of Technology.

All of these initiatives are generating useful information on pricing models and the economics of digital preservation. The LOCKSS networks are especially interesting from the point of view of economic alignment and economic sustainability. The North American networks in particular—the MetaArchive Cooperative, ADPNet, and COPPUL—are financially self-sustaining and have devised membership fee schedules that are affordable for smaller, poorly resourced institutions. Taken together, their experience suggests that using open-source software, aiming for lightweight administrative structures, and delegating as much decision-making power as possible to the member institutions contribute to economic sustainability and can promote economic alignment among otherwise very different networks. Whichever approach or solution one chooses, however, it is advisable to keep it as simple and inexpensive as possible. Simplicity contributes to economic sustainability; complexity undermines it. This maxim rings true across a whole spectrum of activity as there is a great deal of anecdotal evidence to suggest that preservation is a hard sell because of the perceived complexity of its processes. It is true that without recourse to technical effort and knowledge a non-specialist will struggle to gracefully embed current preservation tools into a local infrastructure, let alone be able to wrestle with the complexities of developing an emulated environment for legacy software to run in. But these issues are a distraction from the fact that all the principle preservation issues, certainly at a managerial level, are almost exclusively non-technical.

What is required is clear and attractive advocacy material that focuses on the issue of what decisions are required *to effectively*

deal with content. At some stage, someone in the institution will have to take responsibility for technology choices but those decisions will be inordinately easier, and will be taken more effectively and probably more economically, if they are working from a clear specification, with clear policy guidelines, and in the context of a considered, responsible, and joined-up set of information policies.

In a risk-averse atmosphere, trust is a valuable commodity and audit and certification of preservation environments and processes can be helpful as assurance for organizations to make investments they would otherwise be nervous of making. The development of standards is a form of assurance and as mentioned above, the OAIS model sits alongside other ISO entities (such as ISO 15489:2001 for records management) to offer a useful formal framework to build on. There are a number of models, such as: the DINI (Deutsche Initiative für Netzwerkinformation = German Initiative for Networked Information) framework and DINI-Certificate (2002); the DANS (Data Archiving and Networked Services) Data Seal of Approval (2005-2006); the DRAMBORA (Digital Repositories Audit Method Based on Risk Assessment) audit tool (2006-2007); and the TRAC (Trustworthy Repositories Audit and Certification) checklist (2007).[12] There is also now an ISO-approved process for preservation certification, the TRAC standard (ISO 16363:2012), which may provide the basis for an audit/assessment option that is both effective and affordable, especially once the audit component can be delivered by a trusted and sustainable agency.[13] Thanks to collaborative work between the TRAC Task Force, the Consultative Committee for Space Data Systems (CCSDS), and the Alliance for Permanent Access (APA), individuals and agencies may soon be able to be certified to provide TRAC assessments. If this comes to pass, it will represent significant progress.

---

[12] See DINI: http://www.dini.de/; DINI-Certificate: http://www.dini.de/dini-zertifikat/; DANS: http://www.dans.knaw.nl/; DRAMBORA: http://www.repositoryaudit.eu/; TRAC: http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories (all last accessed 03-08-2012).

[13] See "Space data and information transfer systems—Audit and certification of trustworthy digital repositories" (ISO 16363:2012): http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510 (last accessed 05-16-2012).

The challenge of reducing complexity and streamlining preservation functions also applies to metadata. The PREMIS Data Dictionary is a comprehensive and exhaustive catalogue of nearly all of the fields that one might need for the purposes of preservation and is one of the standard works of reference in the field.[14] Combined with various work that examined the significant properties of information (e.g. the CEDARS and INSPECT projects in the United Kingdom[15]), and work on representation information (carried out in the context of the PLANETS and CASPAR projects amongst others), there has been a great deal of progress made in understanding what technical, descriptive, and administrative data may be required to effectively describe digital material for long-term preservation purposes.

But a gap remains between understanding the ideal metadata requirements for digital objects and choosing to implement that metadata in real-world situations. That gap has to do with time and resources and is therefore an economic issue. Metadata is currently laborious to comprehensively and effectively assign to digital objects in a manner that is likely to satisfy all of their potential future use scenarios. Manual tasks, or even semi-automated tasks, of this nature will not scale up to the level that most organizations require. Whilst technical metadata extractors such as DROID, JHOVE, and the NLNZ Metadata Extractor can harvest useful information, work is still required to work out ways of either automatically extracting or intelligently tagging objects such that they align with institutional policies around value and retention. Automated ways of managing digital objects require machine-readable protocols, which in turn require reliably and persistently identified components. There are different views on the best identifier sets for all sorts of purposes, including file formats, subject classification terms, organizational identity, researcher identity, and so on and so forth, but the overarching issue once again is one of trust—which in turn often depends on prospects for sustainability, which ultimately leads back to economics.

---

[14] See PREMIS Data Dictionary: http://www.loc.gov/standards/premis/ (last accessed 03-08-2012).

[15] See The CEDARS Project: http://www.ukoln.ac.uk/services/elib/projects/cedars/; and INSPECT: http://www.significantproperties.org.uk/ (both last accessed 03-08-2012).

Learning from failed initiatives

It is important to build on success in designing economically sustainable digital preservation programs. It is equally important to learn from unsuccessful initiatives. For example, the Preserving Access to Digital Information (PADI) project was a digital preservation subject gateway set up and maintained by the National Library of Australia (NLA) from the mid-1990s until late 2010. The project was discontinued in that year, primarily because of business decisions about resourcing. "Subject interest, expertise and enthusiasm are vitally important but not sufficient," one of the project participants later observed. "Ongoing sustainability of a service like PADI over a long period probably also requires some dedicated discretionary budget funds, not just a few dedicated individuals. It also requires some available expertise in the means of communication, not just the content." Another important element contributing to sustainability is sharing the "ownership" of a program among a number of institutions and building community engagement in it, even at the expense of managerial efficiency. Again, the fate of the PADI project illustrates the dangers of concentrating ownership in one institution: "The other significant development that came with, and contributed to the growing success of PADI as an information gateway, was a local decision against collaboration, taking control of PADI away from a diverse committee of organizations, and investing it in one institution.[...] A case of making it much more easy to manage, but closing off local commitment to its survival and usefulness."[16] This lesson has been absorbed by the Private LOCKSS Networks in North America, whose governance policies were designed to ensure that management of the networks is shared by or rotates among the participating institutions, thereby building a sense of shared ownership.

Services and (more) business models

As stated in various ways previously, preservation is not an activity that easily lends itself to being configured for delivery as a business practice or commercial enterprise. One of the conclusions of the BRTF-SDPA Final Report (2010, p. 1) argued that devising strategies for preserving digital materials was made difficult by four inherent factors:

---

[16] Personal communication between Maurizio Lunghi and Colin Webb, January 2011.

- Long time horizons;

- Diffused stakeholders;

- Misaligned or weak incentives; and

- Lack of clarity about roles and responsibilities among stakeholders

This may explain why the demand for preservation services is still relatively weak, and consequently why the list of commercial vendors offering to supply those services is still fairly short. The United Kingdom-based technology company Tessella has had success, principally (in this area) with its Safety Deposit Box (SDB) system, which is in use in major national archives around the world and has recently been implemented by the Church of Jesus Christ of Latter-Day Saints to deal with their ingest challenge for the Family Search archive. OCLC launched its Digital Archive Service in 2001 and has been marketing it to state libraries and archives, especially those that are already using CONTENTdm, another OCLC product, to manage their digital collections. Ex Libris has a digital preservation product called Rosetta and is building up its customer base. Ex Libris is pursuing an interesting collaboration model with the National Library of New Zealand, which takes the view that working with a commercial vendor offers the best chance for creating and sustaining some of the core services that a preservation system will require, including a file format registry that will sit at the heart of the product and supply an identification function.

It is clear that, if handled in the right way and set up as a mutually beneficial partnership, relationships between vendors and public-sector institutions can bring enormous benefits to client organizations in terms of economic efficiencies and clarity of business processes. There is a strong argument for saying that organizations should play to their strengths. Taking a slightly different approach, it is possible to engage with technology but only on terms that are advantageous to one's own organization. In telecommunications, banking, health-care and most other sectors of society, organizations set out their principles and mission; and then establish their rights, values and basic rules. They then define the components, functionalities, workflow, and models and terms of specific services, and invite competitive tenders to bid for aspects of the work. An example from Italy is the *Magazzini digitali* ("Digital Stacks") project, in which the Ministry of Culture

set up the global architecture and functions of a trusted digital depository (complete with ingest rules and selection criteria for long term preservation) and then put out a call to tender (or, in American usage, a Request for Proposal or RFP) to private companies.[17] Similarly, Auburn University has outsourced the actual digitization of large analogue collections to external vendors. When it comes to digital preservation, however, the librarians at Auburn have been reluctant to entrust such a crucial part of the library's mission to an external vendor, taking the view that the primary responsibility for ensuring the long-term preservation of the human record in digital form ought to rest with public institutions or alliances of public institutions. That view is shared by many other research universities in the United States.[18] That said, there may be room to explore the topic with some major commercial players in the digitization field (e.g. Google). Conversations of this nature being undertaken by the HathiTrust partnership and the Digital Public Library of America (DPLA) initiative at the Berkman Center for Internet & Society at Harvard University will be worth monitoring.

In common with the broadly accepted view that preservation is an international concern and should be tackled using broadly collaborative working methods, preservation is also increasingly being viewed as a process and a workflow that need not be dealt with by an end-to-end local process. The cost efficiencies and the accelerated development processes that accompany collaborative working can enhance the preservation workflow and can relieve institutions of training and technical overheads that they may not be equipped to meet.

Disaggregated services for preservation were much in vogue several years ago (service-oriented architectures), but the focus has now moved to the potential for cloud services to offer preservation and curation capacity using elastic storage and computing provision. "Trust" remains an issue for organizations contemplating cloud services and whilst one could imagine most services, e.g., replication, hashing, identification, characterization, validation, ingest, migration, verification, authentication, etc., being offered as some form of service, these would need to be

---

[17] Magazzini digitali: http://www.rinascimento-digitale.it/projects-digitalstacks.phtml (last accessed 03-08-2012).
[18] See for example the emerging Digital Preservation Network (DPN) initiative.

underpinned by the type of trusted certification processes mentioned previously (e.g. TRAC, DINI-Certificate).

When faced with hard economic choices about service provision, organizations may inevitably run through a fairly universal set of questions:

- Is this something that we really need?

- How much will it cost?

- How much money have we got?

- How much of what we've got do we want to spend on this?

- Can we get someone else to pay for it?

And in the unlucky event that the answers to all those questions are unsatisfactory, the final question becomes:

- How can we adapt what we already have to do what we need to do?

This is a somewhat long-winded way of illustrating that most organizations are generally forced to make very pragmatic decisions, but in terms of gaps and challenges, it follows that the clearer the arguments are for the value of digital materials, the easier it will be to win the argument about funding. This is true irrespective of whether the chosen solution is an entirely outsourced approach (let's pay someone else to do this for us), or an entirely self-managed one (let's do this ourselves, or with a group of like-minded institutions). In either and all cases, the goal should be the same: to codify long-term digital preservation in institutional (or consortial, or national) policy, and incorporate it into an institution's regular workflow.

If the ideal is to embed digital preservation in the core institutional function so firmly that it becomes a line item in the institution's operating budget, then there is also a pressing need to acknowledge and understand all of the steps short of that ideal. Practically speaking, all organizations (except those for whom preservation is the core mission) are probably going to find themselves somewhere down the rungs of that particular ladder for the immediate future. There remain large challenges and gaps in both defining the business case and the business models for preservation but interesting work has emerged in recent years to start classifying and examining possible options. The BRTF-SDPA Final Report (2010, p. 45) lists five "common funding models for

digital preservation." Ithaka S+R has also done some very useful work in producing case studies on sustaining digital resources.[19]

Roles and Responsibilities

When considering what we might refer to as the preservation interrogatives: the "who," "what," "where," "when," and "how" of digital preservation, the question "who" is probably the most interesting (and sometimes the most intractable) question for many people, focusing as it does on the human aspect and drilling down into the detail of who is actually responsible for preserving material.

There is currently uncertainty within many institutions about who ought to take responsibility for the long-term stewardship of digital content. This is also reflected at the macro-level, where funding bodies, government agencies, institutions and individuals are looking from one to the other trying to work out their moral and financial responsibilities *vis-à-vis* content that is of interest to them.

In terms of building *capability* to preserve, this could be characterized as a problem that funders are interested in. Organizations such as JISC, the Library of Congress, the Institute of Museum and Library Services (IMLS), the Deutsche Forschungsgemeinschaft (DFG), the SURF Foundation in the Netherlands, and various other agencies that support research and innovation have a vested interest in ensuring that the communities that they support have the tools and frameworks and infrastructure that they need to manage the information that they produce in the course of their largely education-related activities. When focusing on the *capacity* to preserve, this is arguably more likely to devolve to institutions and organizations whose responsibility it is to deal with the logistics of having staff in place with the right skills to do the work that the institution requires of them.

When it comes to the *sustainability* of both of the above, then this is where the gap or challenge becomes identifiable. A funder may be able to commission the creation of a useful tool or resource but ensuring that the transfer of that capability into the institution

---

[19] See for example Maron, Nancy L., Kirby Smith, K., & Loy, Matthew. (2009). Sustaining Digital Resources: An On-the-Ground View of Projects Today, *Ithaka Case Studies in Sustainability, Strategic Content Alliance*. Available at: http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability (last accessed 03-08-2012).

actually happens is an uncertain proposition, particularly given that short-term "soft" funding often results in the loss of staff at the end of a project, (thereby also affecting the organization's capacity to preserve). But these are not extraordinary problems. Staff members come and go all the time. Perhaps the answer to this sustainability challenge lies with the sorts of membership bodies that are formed to represent and support different types of organizations. For example, the Digital Preservation Coalition (DPC), the Open Planets Foundation (OPF), and the Alliance for Permanent Access (APA) are all designed to support the transfer of knowledge within and beyond the different domains of activity where digital preservation is a live issue. By the coordination of activity in (and between) areas such as science, humanities, publishing, archives, museums, libraries, galleries, government, etc., it should be possible to establish a more effective collective understanding of how information professionals working in a great variety of different contexts can more effectively preserve digital materials. In economic terms, the issue revolves around how to ensure that the benefits of membership justify the cost of joining.

As already stated above, preservation issues for the majority of people revolve around non-technical issues and when focusing on roles and responsibilities in this domain, the discussion at some point needs to drill down to a granular level, and ultimately requires someone to take some form of position on the nature and the value of the content in question. In any discussion of the economics of preservation, "value" is an interesting word: different from "cost;" and not as practical as "benefit." But if we can establish who regards the content as valuable, then we may arrive at a better understanding of who the potential beneficiaries of the preservation process are. We may then be able to find out if anyone is likely to benefit from that preserved content without contributing to the cost of its preservation, which is of course a ubiquitous scenario in a digital world where instant global access to a dazzling universe of material has become not only common but expected. This is what the BRTF-SDPA Final Report (2010, p. 45) (and the language of economics) calls the "free rider" problem.

In some contexts, universal permanent access is not only a convenient by-product of digitized material finding its way into an open preservation environment, but is the intended and funded outcome. Legal deposit arrangements in various countries are the logical conclusion to the information as "public goods" arrangement, where taxes pay for comprehensive tranches of

material to be made available in perpetuity (sometimes under particular access conditions) by trusted public repositories. But in many other contexts and for the vast majority of institutions and organizations, this is an irrelevance. These bodies have budgets to balance and priorities to define and are very conscious of the opportunity costs of assigning precious resources to an enterprise as currently ill-defined as long-term preservation. At some point, the question will be asked, "Who is going to pay for this?" Should the creator pay? Should the user pay? Should responsibility fall to the institution? Or is it a public problem?

Perhaps one way to examine this problem is to take a step back and look at the creation or acquisition process and work through the decisions that are involved at the instigation of this whole process. In some instances, the case for acquiring a digital file is straightforward. Where the original analogue object is unique or at risk, there is a clear justification for creating a surrogate and this also indicates ownership and interest in the digital file. As a surrogate, the physical object and its digital manifestation are related. In cases where a physical object needs to be copiously used by a great variety of people, there is also a clear justification for digitization, although given that the original is probably sturdy and common, the subsequent stewardship issues begin to get murky when questions are asked about the point of storing something that can be easily accessed in a number of other ways.

The following represent four selection criteria elements that might help inform policy-making:

- Are we allowed to preserve it? (Who owns it?)

- Is there someone (right now) that wants to use it?

- Can we carry on making it accessible? (Will it be technically possible?)

- How interesting is the information? (Will someone want it in the future?)

As stated before, selection is an absolutely key part of effective preservation practice, particularly as we exist in a period where analogue material is likely to be with us for some time to come whilst the amount of new digital material requiring storage grows all the time.

It may be possible for some organizations to settle on fairly loose or general policies towards responsibility for material, such as forming the view that any decision to ingest material into a given preservation environment implies the acceptance of responsibility, and therefore the acceptance of ongoing cost. Other general statements of this nature may be applicable also, but there is a potential problem with this approach in that the stewardship of digital material and collections is not a static and tidy problem. As digital objects progress through a lifecycle, their value—like any investment—may rise and fall. Perhaps what is needed is some low-overhead administrative (or even just conceptual) way to keep track of three vital pieces of information that will assist content owners with the ongoing challenge of appraisal, which can be defined as the iterative selection process that ideally takes place at various points subsequent to the initial selection decision.

The role of *creator* of the digital object/collection/dataset is fairly clear and should often be reflected in the metadata associated with an object, or will be known to those managing the environment that the object is destined to be stored in. This is often a key piece of information for a great variety of reasons but may also be important for appraisal purposes. What is less obvious, and not by any means likely to be the same as the *creator* is the identity of a person who might be referred to as the *principal keeper*. This would refer to someone who has appropriate authority and is interested in knowing that the object(s) in question are supposed to be residing in the preservation environment. The third piece of information that might be useful to know is who the *principal user* is. This would refer to someone who had self-identified themselves as a person who was interested in the object(s) in question and who had a vested interest in seeing that they continued to be stored safely.

In many environments, one suspects that these designations would not make much sense as two, or perhaps all three, of the designations would be the same person. But in other cases—particularly perhaps where special collections of digital material were stored for long periods of time (at some expense) and the original motivations for archiving the material had become unclear—designations of this type may be helpful in determining ongoing value.

In order for this proposal to be valuable, refinements would need to be introduced whereby the identity of the keeper or the user would be passed on as appropriate to new incumbents or to

others taking on the mantle of research or teaching in that area (if that was the use case). An action would be triggered however if at any point the keeper or the user identity changed or became blank—that is, if one or other of those roles became vacant in relation to an object. This would alert the host organization to the fact that either somebody thought that the object was no longer worth storing, or that the file was no longer worth using, either of which represents a strong case for disposal. A number of other refinements (e.g. designated community alerts and automatic retention periods) could be introduced as safeguards but the point would be to try and tackle the problem of unmanaged persistence.

What Success Looks Like: A Five-Year Forecast

One measure of success will be that analogue and digital documents are considered and treated equally in any preservation regime. The current practice of acquiring, cataloguing, protecting, and making available predominantly or even only analogue materials while postponing similar treatment for digital content entails possibly irrecoverable losses to the corpus of cultural heritage materials and important research resources.

Of course, given limited resources, selection and prioritization will have to be applied to both types of resources. This will require fundamental changes in strategic planning and organization at many institutions. Cultural heritage institutions are by nature conservative, and transforming them will be far from easy. However, the authors agree that these changes will be necessary to achieve success in this field. When normal practice within an organization automatically factors in the whole lifecycle costs of acquiring or creating a digital collection (including the opportunity costs) and the institution has a clear view of the likely short, medium, and long-term benefits of doing so, then it might be possible to claim that the role of digital preservation is as innately understood within an organization as (analogue) archival practice or records management. Fortunately, as the body of this paper shows, there are a growing number of successful transformations underway. Taken together, they suggest that momentum is building in the right direction.

In that connection, success in this area will begin with an institutional recognition that long-term digital preservation is a high-priority activity that requires an ongoing commitment of time and resources. This will involve having policies that are broadly meaningful across institutions and model governance instruments

that can be adapted to reflect local conditions and practices. It will mean that staff members are trained in basic preservation competencies, ranging from digitization best practices and optimal file organization to writing instructions for digital preservation software (e.g. LOCKSS manifests and plugins). And it will mean that digital preservation is embedded into the institutional way of behaving and operating (i.e. linked to policy and workflow measures) and embodied in an optimum balanced budget from the start.

In short, we will have achieved success when long-term digital preservation becomes a routine and economically sustainable activity and a generally accepted part of the mission of cultural heritage organizations and other stakeholders—that is, when most institutions have incorporated the long-term stewardship of digital materials into their day-to-day operations, preferably with some degree of mutual assistance and coordination. This may happen as a result of national policy and government mandates, or because of a series of local initiatives. The main thing is that it happens—and in a sustainable way, with long-term institutional commitment, public understanding and support, budget lines, and dedicated personnel.

To that end, the authors propose the following guiding principles:

- Digital preservation should be an integral part of all of projects dealing with the digitization of analogue documents and/or the acquisition of born-digital documents having to do with the national cultural heritage.

- Digital preservation is not a luxury. Ensuring adequate protection for digital content should be just as much a part of an institution's workflow as protecting analogue materials from water, fire, or careless handling.

- More broadly, digital preservation should also figure in national public policy. Recipients of public funding (libraries, museums, archives, etc.) should be required to include digital preservation in their activities, build and share a knowledge base, and pool resources to develop or add to preservation tools and services. The recent requirement by the National Science Foundation (NSF) in the United States that grant

applicants submit a long-term data management plan is just one example of this.[20]

- Sufficient funding should be dedicated exclusively to digital preservation. Large-scale publicly funded digitization initiatives that do not also include a budget and a clearly defined strategy for digital preservation are disasters waiting to happen and an unwise use of public monies.

As was pointed out at the beginning of this paper, digital preservation is a relatively new area of activity for most cultural heritage organizations. It is all the more important, therefore, to share experiences, tools, and successful approaches across institutions and countries.

Towards Economic Alignment: Ten Recommendations

The following set of ten recommendations is intended to address economic and cost issues and to promote economic alignment among digital preservation initiatives in different countries. It reflects the cumulative experience of the authors and incorporates discussion points that arose at the ANADP conference.

- Develop and launch a coordinated international campaign to make Archive/Library/Museum (ALM) directors and administrators aware that long-term digital preservation requires stable funding and a continuous allocation of resources. ALMs and scientific institutions need specific, practical suggestions for incorporating digital preservation into their budgets. Some of this work is already being done by national libraries and archives in individual countries (e.g. the National Digital Stewardship Alliance in the United States). These efforts need to be coordinated.

- Establish a Digital Preservation Resource Centre (DPRC). Decision-makers at ALMs need a single place where they can find current information on various digital preservation solutions. Ideally, this resource centre—which we are provisionally calling a Digital Preservation Resource Centre (DPRC)—should address three key areas: awareness, tools,

---

[20] See NSF Data Management Plan Requirements:
http://www.nsf.gov/eng/general/dmp.jsp (last accessed 03-08-2012).

and hosting. It should include case studies (including best practices as well as failed initiatives), data from benchmarking exercises, and technical evaluations of systems performance. It should also contain information on a palette of economic approaches and solutions, ranging from proprietary commercial and vendor solutions (e.g. Ex Libris Rosetta, or Tessella SDB, the OCLC Digital Archive, Portico) to community-owned, member-managed solutions (e.g. the HathiTrust or the MetaArchive Cooperative). These solutions could be arranged by format, with transparency about costs, rights, and responsibilities being essential. In designing the portal, we can take a lesson from the IT industry. In the late 1980s, as the market for desktop workstations and enterprise servers was taking off, a small number of workstation vendors formed the System Performance Evaluation Corporation (SPEC). Based in Gainesville, Virginia, SPEC defines its goal as "ensur[ing] that the marketplace has a fair and useful set of metrics to differentiate candidate systems"[21] by providing standardized source code based on existing applications that can be used in benchmarking exercises. Another possible model is The Keepers Registry based at the University of Edinburgh. This is currently a registry of e-journal preservation services but could be developed further to address issues related to metrics. Questions for further discussion include the level of detail, openness, and transparency (e.g. whether the portal should include specific information about failed preservation efforts or the neglect or loss of materials), as well as funding and sustainability.

- Share experience, objectives, tools, documentation (including governance policies), and practices with other preservation initiatives and communities. The additional effort and cost of doing so should be understood as a prudent investment in the sustainability of digital preservation in general. Given the growing body of successful experience, no institution, consortium, or country should have to navigate the challenges of digital preservation in isolation. Similarly, every institution should be prepared to contribute knowledge and experience back to the general preservation community. Specific recommendations for promoting partnerships and cooperation

---

[21] System Performance Evaluation Corporation (SPEC): http://www.spec.org/ (last accessed 03-08-2012).

include periodic conferences like ANADP and other events at the national or international level (e.g. iPres); the aforementioned Digital Preservation Resource Centre (DPRC); and Distributed Preservation Development Networks (DPDNs). The DPRC could include a "technology watch" section and a brokerage service for open-source developers and users to share experiences and solutions.

- Assemble and make available case studies of digital preservation costs. Although costs cannot be predicted with certainty, benchmark figures and real-life cost scenarios are useful. Case studies of cost and business models are emerging in particular from some of the projects funded by JISC, which is committed to supporting research on cost issues and making this information and the methods of organizing and obtaining it as widely available as possible.

- Develop a matrix of selection criteria for digital preservation—in other words, a digital-preservation "triage chart." Digital content is easy to produce. Preserving it can be complex and expensive. For this reason, ALMs must decide what they want to preserve, why, for how long, and for what level of use. Appraisal and selection must reflect user requirements (both actual and anticipated) and legal constraints, if any. As part of this effort, the community should compile a list of selection best practices for specific types of institutions, types of content, and user communities.

- Study and (where appropriate) promote community-owned solutions. Community-owned digital-preservation initiatives are gaining currency and credibility. For example, the MetaArchive Cooperative is an international LOCKSS-based network with a good track record and relatively low barriers to entry. The same could be said of ADPNet, COPPUL, and other community-owned networks in North America. HathiTrust is another interesting example of an international community-based partnership in action. Initiatives like these enable practitioners to pool resources and share experience. That said, the community still needs viable business models to create a financially sound digital-preservation development community (e.g. JHOVE and JHOVE2). The Open Planets Foundation (OPF) may be a possible model for this. The community also needs mechanisms for billing, hosting, and assigning prices to digital-preservation products and services. Here it is important to recognize that "sweat equity" (i.e. in-

kind contributions by member institutions) can be a useful currency. The OPF relies on this model: the charter members pay for the administration and organizational costs and the associate members provide the "sweat." The authors propose setting up a brokerage mechanism, perhaps in the form of a registry of developers who are willing to trade expertise with others through the aforementioned Distributed Preservation Development Networks (DPDNs). Skillshare—a Web-based teaching and learning exchange—could be a possible model for the DPDN brokerage.

- Explore opportunities for public-private partnerships. Public institutions and private businesses have very different missions and priorities, but there may be areas in which they can cooperate in mutually beneficial ways. Google Books and HathiTrust in the United States are two examples of apparently successful public-private partnerships; *Maggazzini digitali* in Italy is another. Building successful partnerships depends on standardizing the preservation needs of public-sector institutions and creating conditions in which private companies can compete to meet those needs against an agreed-upon set of benchmarking criteria. It also depends on persuading private companies to participate in preserving society's patrimony and cultural heritage, perhaps through public recognition or even preferential fiscal (read: tax) policies. The BRTF-SDPA Final Report identified incentives and business models for public-private cooperation, but the solutions tend to be country-specific and the state of research in this area is still undeveloped. The community needs to identify other activities and suggest new initiatives to tackle the topic of public-private partnerships. The current large-scale EU-funded initiatives in whcih a range of organizations (including commercial partners) are looking at preservation issues might serve as a good starting-point.

- Add digital preservation to the library-school curriculum. Adding a standardized course on digital preservation to the curriculum and investing in post-graduate professional development in digital preservation are good ways to inculcate an understanding of the economics of digital preservation and promote international alignment in this area. Training programs in digital preservation should focus on common technologies and standards and should culminate in the awarding of an international certificate in digital preservation.

This will help to facilitate cooperation in this area by inculcating a common understanding of key concepts and a common skill set.

- Define core services. We have argued that a clear definition of roles and responsibilities is crucial for digital preservation. The same thing goes for core services, an area in which we need to take our cue from the larger user community. We should look to the user community to identify key services, coordinate initiatives, promote common standards, implement policies and recommendations, and encourage the use of basic services like Trusted Digital Repositories (TDR) and Persistent Identifiers (PI) for preservation networks and preserved materials. It would be strategically useful to "standardize" some key services across user communities in order to offer tested, universally applicable solutions for end-users and to stimulate competition among technology providers, which should in turn lead to lower prices. Certification tools for trusted digital repositories include TRAC in the United States and DINI in Germany; DOI, Handle, NBN are examples of protocols for persistent identifiers.

- Support research and development. Finally, encourage research and development (R&D) in digital preservation in order to identify tools and services that yield the best return on investment. This is an area in which external support from government or private funding agencies can play a useful— indeed, a crucial—role, and ALMs should work together across national boundaries to identify and apply for suitable opportunities. Research and development on various aspects of digital preservation could also be added to the curricula at schools of library and information science in North America, the United Kingdom, Europe, and around the globe. Indeed, we need to move beyond our focus on North America and Europe and make connections with digital preservation initiatives in Latin America, Africa, Asia, Australia, and New Zealand.

Conclusions

In a 2004 article whose title was inspired by American poet Wallace Stevens' "Thirteen Ways of Looking at a Blackbird," Brian Lavoie and Lorcan Dempsey recognized that digital preservation is "an economic process, in the sense of matching

limited means with ambitious objectives." They were right on both counts: the means are limited and the objectives are indeed ambitious. As this paper shows, however, an impressive—one might even say "ambitious"—amount of work has already been done in Europe, North America, and elsewhere on identifying the costs of digital preservation and devising tools, techniques, and procedures for absorbing those costs into ongoing preservation programs. Moreover, this work has been accomplished in large part by realizing economies through collaboration among institutions. Despite their different origins, missions, and management structures, the preservation initiatives identified in the body of this paper: the *Maggazzini digitali* project in Italy; HathiTrust, the MetaArchive Cooperative, and the Alabama Digital Preservation Network in the United States; the COPPUL PLN in Canada; the Digital Curation Centre in the United Kingdom; the Open Planets Foundation; and so forth—prove that it is possible to take advantage of accumulated experience and community-based effort to build working, economically sustainable digital preservation networks across states, provinces, and even countries. In Lavoie's and Dempsey's (2004) words, digital preservation "is an ongoing, long-term commitment, often shared, and cooperatively met, by many stakeholders." The task facing us now is to build on the collaborative work that has been done.

References

AIDA: http://aida.jiscinvolve.org/wp/ (last accessed 06-06-2012).

Alabama Digital Preservation Network (ADPNet): http://adpn.org/ (last accessed 06-06-2012).

Alliance for Permanent Access (APA): http://www.alliancepermanentaccess.org/ (last accessed 06-06-2012).

Archivematica: http://www.archivematica.org/wiki/index.php?title=Main_Page (last accessed 06-06-2012).

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2008). *Sustainable 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. Retrieved July 29, 2011 from

http://brtf.sdsc.edu/biblio/BRTF_Interim Report.pdf (last accessed 06-06-2012).

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information.* Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. Retrieved July 29, 2011 from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (last accessed 06-06-2012).

Borgman, C.L. (2007). Scholarship in the digital age: Information, infrastructure, and the Internet. Cambridge, MA: MIT Press.

Borgman, C.L., Wallis, J.C., & Enyedy, N. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. *10th European Conference on Digital Libraries. Alicante, Spain*: Berlin: Springer.

California Digital Library (CDL): http://www.cdlib.org/ (last accessed 06-06-2012).

Chronopolis: https://chronopolis.sdsc.edu/ (last accessed 06-06-2012).

Council of Prairie and Pacific University Libraries (COPPUL) Private LOCKSS Network: http://www.coppul.ca/pln.htmlEconomic-v2.docx (last accessed 06-06-2012).

Curator's Workbench: http://www.lib.unc.edu/blogs/cdr/index.php/2010/12/01/announcing-the-curators-workbench/ (last accessed 06-06-2012).

Currall, J., & McKinney, P. (2006). Investing in Value: A Perspective on Digital Preservation. *D-Lib Magazine* 12 (4). http://www.dlib.org/dlib/april06/mckinney/04mckinney.html (last accessed 06-06-2012).

Data Asset Framework (DAF): http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework (last accessed 06-06-2012).

Deutsche Initiative für Netzwerkinformation: http://www.dini.de/
(last accessed 06-06-2012).

DICE Center:
http://www.diceresearch.org/DICE_Site/iRODS_Uses.ht
ml (last accessed 03-08-2012).

Digital Curation Centre. (2005). *Digital curation and preservation:*
Defining the research agenda for the next decade. In
Report of the Warwick Workshop, November 7-8, 2005.
Digital Curation Centre: Warwick, UK. Retrieved July 27,
2007, from
http://www.dcc.ac.uk/events/warwick_2005/Warwick_W
orkshop_report.pdf (last accessed 06-06-2012).

Digital Preservation Coalition (DPC): http://www.dpconline.org/
(last accessed 06-06-2012).

Digital Public Library of America (DPLA): http://dp.la/ (last
accessed 06-06-2012).

DRAMBORA: http://www.dcc.ac.uk/resources/tools-and-
applications/drambora (last accessed 06-06-2012).

DROID and PRONOM:
http://www.nationalarchives.gov.uk/PRONOM/Default.as
px (last accessed 06-06-2012).

DuraCloud: http://www.duracloud.org/ (last accessed 06-06-2012).

Esanu, J., Davidson, J., Ross, S., & Anderson, W. (2004).
Selection, appraisal, and retention of digital scientific
data: Highlights of an ERPANET/CODATA workshop.
Data Science Journal 2004, 3. Retrieved July 30, 2007,
from http://www.jstage.jst.go.jp/article/dsj/3/0/227/_pdf
(last accessed 06-06-2012).

Effective Strategic model for the Preservation and disposal of
Institutional Digital Assets (espida):
http://www.gla.ac.uk/espida/ (last accessed 06-06-2012).

Ex Libris Rosetta:
http://www.exlibrisgroup.com/category/RosettaOverview
(last accessed 06-06-2012).

Garrett, J. and Waters, D. (1996). *Preserving Digital Information:*
Report of the Task Force on Archiving of Digital
Information. Washington, D.C.: Commission on
Preservation and Access; Research Libraries Group.

Retrieved July 29, 2011, from
http://www.clir.org/pubs/reports/pub63watersgarrett.pdf
(last accessed 06-06-2012).

HathiTrust: http://www.hathitrust.org/ (last accessed 06-06-2012).

HOPPLA: http://www.ifs.tuwien.ac.at/dp/hoppla/ (last accessed
06-06-2012).

Internet Archive: http://www.archive.org/ (last accessed 03-08-
2012).

ISO/IEC Guide 73. (2002). Risk management. Vocabulary.
Guidelines for use in standards.

JHOVE: http://hul.harvard.edu/jhove/ (last accessed 06-06-2012).

The Keepers Registry: http://thekeepers.org/thekeepers/keepers.asp
(last accessed 06-06-2012).

Keeping Research Data Safe (KRDS):
http://www.beagrie.com/krds.php (last accessed 06-06-
2012).

Lavoie, B., and & Dempsey, L. (2004). Thirteen ways of looking
at: digital preservation. *D-Lib Magazine* 10 (7/8).
doi:10.1045/july2004-lavoie (last accessed 06-06-2012).

Lifecycle Information for E-Literature (LIFE):
http://www.life.ac.uk/ (last accessed 06-06-2012).

LOCKSS und KOPAL Infrastruktur und Interoperabilität:
http://www.lukii.hu-berlin.de/ (last accessed 06-06-2012).

Maggazzini digitali (Digital Stacks): http://www.rinascimento-
digitale.it/projects-digitalstacks.phtml (last accessed 06-
06-2012).

Maron, N.L., Kirby Smith, K., & Loy, M. (2009). Sustaining
Digital Resources: An On-the-Ground View of Projects
Today, Ithaka Case Studies in Sustainability, Strategic
Content Alliance. Retrieved April 28, 2010, from
http://www.ithaka.org/ithaka-s-r/research/ithaka-case-
studies-in-sustainability (last accessed 06-06-2012).

Mayer-Schönberger, V. (2009). *Delete: The Virtue of Forgetting in
the Digital Age.* Princeton, NJ: Princeton University
Press.

Mayernik, M.S., Wallis, J.C., Borgman, C.L., & Pepe, A. (2007).

Adding context to content: The CENS Deployment Center. *Proceedings of the American Society for Information Science and Technology* 44 (1). Milwaukee, WI: Information Today. doi:10.1002/meet.1450440388 (last accessed 06-06-2012).

MetaArchive Cooperative: http://www.metaarchive.org/ (last accessed 06-06-2012).

National Digital Information Infrastructure and Preservation Program (NDIIPP): http://www.digitalpreservation.gov/ (last accessed 06-06-2012).

National Digital Stewardship Alliance (NDSA): http://www.digitalpreservation.gov/ndsa/ (last accessed 06-06-2012).

National Library of New Zealand (NLNZ) Metadata Extraction Tool: http://www.natlib.govt.nz/services/get-advice/digital-libraries/metadata-extraction-tool (last accessed 06-06-2012).

National Science Foundation. (2003). *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Retrieved July 29, 2011, from http://www.nsf.gov/publications/pub_summ.jsp?ods_key =cise051203 (last accessed 06-06-2012).

National Science Foundation Data Management Plan Requirements: http://www.nsf.gov/eng/general/dmp.jsp (last accessed 06-06-2012).

Nielsen, A.B. & Kejser, U.B. (2008). Using the LIFE costing model: case studies from DK. Retrieved July 29, 2011, from http://discovery.ucl.ac.uk/9313/ (last accessed 06-06-2012).

OCLC and CRL. (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Retrieved July 29, 2011, from http://www.crl.edu/sites/default/files/attachments/pages/tr ac_0.pdf (last accessed 06-06-2012).

OCLC Digital Archive: http://www.oclc.org/digitalarchive/ (last accessed 06-06-2012).

Open Planets Foundation (OPF): http://www.openplanetsfoundation.org/ (last accessed 06-

06-2012).

Payette, S., Staples, T., & Wayland, R. (2003, April). The Fedora
        Project: An Open-source digital object repository
        management system. *D-Lib Magazine* 9 (4). Retrieved
        September 30, 2006, from
        http://www.dlib.org/dlib/april03/staples/04staples.html
        (last accessed 06-06-2012).

PADI Archive: http://pandora.nla.gov.au/tep/10691 (last accessed
        03-08-2012).

Portico: http://www.portico.org/digital-preservation/ (last accessed
        06-06-2012).

Preserving Access to Digital Information (PADI):
        http://pandora.nla.gov.au/tep/10691 (last accessed 06-06-
        2012).

Rusbridge, C. (2008, July, 29) Re: "Digital preservation" term
        considered harmful? [Web log message from *Digital
        Curation Blog]*. Retrieved September 7, 2009, from
        http://digitalcuration.blogspot.com/2008/07/digital-
        preservation-term-considered.html (last accessed 06-06-
        2012).

Santini, M. (2004a). A shallow approach to syntactic feature
        extraction for genre classification. Proceedings of the 7th
        Annual Colloquium of the UK Special Interest Group for
        Computational Linguistics.

Santini, M. (2004b). *State-of-the-art on automatic genre
        identification*. (Technical Report ITRI-04-03). University
        of Brighton, UK, Information Technology Research
        Institute (ITRI).

Skillshare: http://www.skillshare.com/learn (last accessed 06-06-
        2012).

Skinner, K., & Schultz, M., eds. (2010). *A Guide to Distributed
        Digital Preservation*. Atlanta, GA: Educopia Institute.
        Retrieved April 28, 2011, from
        http://www.metaarchive.org/GDDP (last accessed 06-06-
        2012).

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G.,
        Stuve, D., et al. (2003, January). DSpace: An open source
        dynamic digital repository. *D-Lib Magazine 9 (1)*.

Retrieved September 30, 2006, from
http://www.dlib.org/dlib/january03/smith/01smith.html
(last accessed 06-06-2012).

System Performance Evaluation Corporation (SPEC):
http://www.spec.org/ (last accessed 06-06-2012).

Tessella SDB Digital Preservation: http://www.digital-
preservation.com/ (last accessed 06-06-2012).

UC3Merritt: http://merritt.cdlib.org/ (last accessed 06-06-2012).

UK LOCKSS Alliance: http://www.lockssalliance.ac.uk/ (last
accessed 06-06-2012).

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine
learning tools and techniques*. (2nd ed.). San Francisco,
CA: Morgan Kaufmann.