http://conference.ifla.org/ifla78 Date submitted: 31 May 2012

Safety in Numbers: Distributed Digital Preservation Networks

Aaron Trehub
Auburn University
Auburn, Alabama, USA
E-mail: trehuaj@auburn.edu

Martin Halbert
University of North Texas
Denton, Texas, USA
E-mail: Martin.Halbert@unt.edu

Session: 216 — *Continuity in the face of digital disasters: Disaster planning and recovery for digital libraries* — Information Technology

Abstract:

It has long been recognized that there is safety in numbers and that redundancy enhances survivability. This principle has been applied in many spheres of human activity, from engineering to military science. It is now being applied in librarianship and digital preservation, through the creation of distributed digital preservation (DDP) networks using the open-source LOCKSS ("Lots of Copies Keep Stuff Safe") software. This paper describes two Private LOCKSS Networks (PLNs) based in North America: the MetaArchive Cooperative, an international preservation network serving more than 50 member institutions in the U.S., Brazil, Spain, and the U.K.; and the Alabama Digital Preservation Network (ADPNet), a state-based preservation network serving academic libraries, public libraries, and the state archives in Alabama. The paper argues that PLNs offer a technologically robust, administratively manageable, and economically sustainable way to protect digital assets and ensure the continuity of digital libraries in the face of natural and man-made disasters.

Introduction

The notion that there is safety in numbers is an old one in human history. Survivability—beating the odds—has long been linked to numbers and redundancy. Medieval scribes made multiple copies of manuscripts. Rural families had many children in order to ensure that there would be enough hands to do the field work and take care of the parents in old age. Redundancy is a time-honored principle in engineering, and bridges, buildings, and airplanes all have backup

systems aimed at preventing catastrophic failure. The early strategists of the nuclear age—the "wizards of Armageddon", in Fred Kaplan's phrase[1]—identified redundancy and geographic dispersion as crucial elements in the survivability of nuclear weapons and the preservation of a credible strategic deterrent. In a more positive vein, the principle of safety-in-numbers applies to the preservation of cultural artifacts and the human record. Increasingly, that means people are creatively applying it to the preservation of *digital* artifacts.

Digital preservation is the corollary to digital collection building. Like many things having to do with infrastructure, it's invisible, unglamorous, and absolutely necessary. Although precise figures are hard to come by, it is generally recognized that most of the world's information is currently being produced in digital form, not as print documents or analogue artifacts. This poses a serious challenge to libraries, archives, museums, and other cultural memory organizations, as well as government agencies. Unlike their analogue counterparts, digital files are inherently susceptible to decay, destruction, and disappearance. Given the vulnerability of digital content to fires, floods, tornadoes, hurricanes, power blackouts, cyber-attacks, and a variety of hardware and software failures, cultural heritage organizations need to start incorporating long-term digital preservation services for locally owned and created digital content into their routine operations, or risk losing that content irrevocably. The head of the British Library's digital-preservation program, Adam Farquhar, recently remarked that "If we're not careful, we will know more about the beginning of the 20th century than the beginning of the 21st century."[2]

A number of countries have recognized the challenge and embarked on ambitious digital preservation programs at the national level. In the United States, the Library of Congress initiated the National Digital Information Infrastructure and Preservation Program (NDIIPP) over ten years ago, and launched the National Digital Stewardship Alliance (NDSA) in 2010.[3] In the United Kingdom, the Digital Curation Centre (DCC) of the Joint Information Systems Committee (JISC) provides a national focus for digital preservation issues.[4] Similar initiatives are underway in many national contexts, including Canada, New Zealand, France, Germany, Italy, and the Netherlands.

Several lessons have already emerged from these initiatives. One of them concerns the importance of collaboration among institutions, states, and even countries—another example of the principle of strength in numbers. Collaboration has been shown to improve cost effectiveness, promote the sharing of tools and expertise, and strengthen community-based stewardship and control of unique digital assets. With numbers comes complexity, however, and comprehensive digital preservation programs inevitably raise difficult technical, administrative, financial, and even legal questions. These questions are not unsolvable, however. Indeed, they are being solved, or successfully addressed, by a number of preservation programs in the United States, Canada, and other countries. There is a growing body of empirical experience that shows that it is possible to build technically and administratively robust digital preservation networks across institutional and geographical borders without compromising those networks' long-term viability through excessive complexity and cost.

Distributed Digital Preservation and LOCKSS

The authors of the final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (henceforth BRTF-SDPA Final Report) have written that "economically sustainable preservation—ensuring the ongoing and efficient allocation of resources to digital preservation—is an urgent societal problem."[5] Proceeding from that assertion, they posited five conditions for economic sustainability:

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
5. Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities.

Fortunately, digital preservation solutions that satisfy most or all of those five conditions have started to emerge in the past several years. One successful approach uses Distributed Digital Preservation (DDP). As its name implies, DDP is based on the idea of distributing copies of digital files to server computers at geographically dispersed locations in order to maximize their chances of surviving a natural or man-made disaster, power failure, or other disruption. DDP networks consist of multiple preservation sites, selected with the following principles in mind:

- Sites preserving the same content should not be within a 75-125-mile radius of one another;
- Preservation sites should be distributed beyond the typical pathways of natural disasters, such as hurricanes, typhoons, and tornadoes;
- Preservation sites should be distributed across different power grids;
- Preservation sites should be under the control of different systems administrators;
- Content preserved in disparate sites should be on live media and should be checked on a regular basis for bit-rot and other issues; and
- Content should be replicated at least three times in accordance with the principles detailed above.[6]

Some digital preservation initiatives are successfully combining DDP with LOCKSS ("Lots Of Copies Keep Stuff Safe") peer-to-peer software in so-called Private LOCKSS Networks (PLNs). LOCKSS was developed and is currently maintained at the Stanford University Libraries.[7] It is ideally suited for use in DDP networks. Originally designed to harvest, cache, and preserve digital copies of journals for academic libraries, LOCKSS is also effective at harvesting, caching, and preserving multiple copies of locally created digital content for cultural memory organizations in general. LOCKSS servers (also called LOCKSS boxes, LOCKSS caches, and LOCKSS nodes) typically perform the following functions:

- They collect content from target Web sites using a Web crawler similar to those used by search engines;
- They continually compare the content they have collected with the same content collected by other LOCKSS boxes, and repair any differences;

- They act as a Web proxy or cache, providing browsers in the library's community with access to the publisher's content or the preserved content as appropriate; and
- They provide a Web-based administrative interface that allows the library staff to target new content for preservation, monitor the state of the content being preserved, and control access to the preserved content.

LOCKSS is open-source software and therefore theoretically available for further development by the open-source community. In practice, however, its design and development have thus far been confined to the LOCKSS team at Stanford.

DDP Networks in North America

Although there are LOCKSS-based DDP networks in Europe (e.g. the UK LOCKSS Alliance and the LuKII project in Germany), most of the Private LOCKSS Networks are currently based in North America.[8] This paper focuses on two of them: the MetaArchive Cooperative, an international preservation network which began in 2003-2004 with support from the U.S. Library of Congress' NDIIPP Program; and the Alabama Digital Preservation Network (ADPNet), a statewide preservation network which began in 2006 with a two-year grant from the Institute of Museum and Library Services (IMLS), a U.S.-based federal funding agency. ADPNet also served as the model for a third LOCKSS-based network in North America: the Council of Prairie and Pacific University Libraries (COPPUL) PLN in western Canada.[9]

The MetaArchive Cooperative (http://metaarchive.org/) is community-owned, community-led network that preserves digital collections for more than 50 member libraries, archives, and other digital memory organizations in four countries. The Cooperative was founded in 2003-2004 to develop a collaborative digital preservation solution for special collections materials, including digitized and born digital collections. Working cooperatively with the Library of Congress through the NDIIPP Program, the founders sought to embed both the knowledge and the technical infrastructure of preservation within MetaArchive's member institutions. They selected the LOCKSS software as a technical framework that matched the Cooperative's principles, and built additional curatorial tools that layer with LOCKSS to promote the curation and preservation of digital special collections, including newspapers, Electronic Theses and Dissertations, photographs, audio, video, and datasets. In doing so, they created a secure, cost-effective repository solution that fosters ownership rather than outsourcing of this core library/archive mission. The Cooperative moved to an open membership model in 2007, and has expanded in five years from a small group of six southeastern academic libraries to an extended community of more than 50 international academic libraries, public libraries, archives, and research centers. In addition to preserving its members' content, the Cooperative engages regularly with other digital preservation groups in research and development work to enable interoperability of the MetaArchive network with other digital preservation approaches (e.g. Chronopolis at the San Diego Supercomputer Center at the University of California San Diego and the CODA suite of microservices at the University of North Texas). The Cooperative is also studying a number of genre-specific curation and preservation issues, including digitized and born-digital newspapers (with NEH support) and Electronic Theses and Dissertations (with IMLS support).

The Alabama Digital Preservation Network (ADPNet: http://adpn.org/) is a statewide digital preservation network that serves cultural heritage organizations in Alabama. ADPNet currently has nine members: the Alabama Department of Archives & History in Montgomery, Auburn University, the Birmingham Public Library, the Huntsville-Madison County Public Library, Spring Hill College in Mobile, Troy University in Troy, the University of Alabama in Tuscaloosa, the University of Alabama in Birmingham, and the University of North Alabama in Florence. Inspired in large part by Auburn University's experience with the MetaArchive Cooperative, the Alabama network began in 2006 with a two-year National Leadership Grant from the Institute of Museum and Library Services (IMLS). The grant provided support for equipment and associated expenses to the seven founding institutions; crucially, it also covered those institutions' annual membership fees in the LOCKSS Alliance for the same period. For their part, the participating institutions split the equipment costs with the IMLS and contributed staff time and other in-house resources to the project. A LOCKSS staff member was assigned to the project to provide technical support and guidance. The IMLS grant ended in September 2008, and ADPNet is now a self-sustaining, member-owned DDP network operating under the auspices of the Network of Alabama Academic Libraries (NAAL), a department of the Alabama Commission on Higher Education in Montgomery.[10] All of the original member institutions have contributed content to the network, which currently contains over 400 archival units totaling over four terabytes. The network plans to harvest several terabytes of new content in 2012-2013, including content from the public libraries in Birmingham and Huntsville.

It is no accident that the first Private LOCKSS Networks in the United States arose in the southeastern part of the country. The southern states tend to be poorer economically than states in other parts of the country. They also are vulnerable to hurricanes, tornadoes, flooding, and other natural disasters, especially on and around the Gulf coast. For example, Alabama has been hit by at least four major hurricanes and many tropical storms in the past decade. In 2005, Hurricane Katrina devastated the coastal communities of Bayou la Batre and Coden and flooded downtown Mobile. The coastal communities are not the only parts of the state that have suffered from natural disasters, however. The interior of the state is vulnerable to tornadoes. In March 2007 a tornado swept through Enterprise, Alabama, destroying a high school and causing nine deaths.[11] In April 2011, a string of powerful tornadoes hit the cities of Tuscaloosa, Birmingham, and Cullman, destroying entire neighborhoods and killing over 250 people.[12] Despite these challenges, Alabama and other states in the region are home to a rich and growing array of digital collections at libraries, archives, and museums. These include Documenting the American South at the University of North Carolina at Chapel Hill, the Digital Library of Georgia at the University of Georgia, and AlabamaMosaic, a statewide repository of digital materials on all aspects of Alabama's history, geography, and cultures, at the Network of Alabama Academic Libraries (NAAL) in Montgomery.[13] This combination of circumstances—extreme weather, meager state financial resources, and rich digital collections—has made the southeastern United States an ideal test case for a simple, inexpensive, but effective digital-preservation solution like LOCKSS.

Disaster Recovery: Planned and Unplanned

LOCKSS-based Distributed Digital Preservation networks are designed to ensure that digital content will survive an array of threats, ranging from natural or man-made disasters to hardware and software failures. Indeed, that is their reason for being. The members of the MetaArchive Cooperative and the Alabama Digital Preservation Network are mindful of this fact, and have over the years designed and performed periodic disaster-recovery exercises to test their networks' robustness. Basically, these exercises fall into two categories: exercises for restoring a damaged or destroyed LOCKSS node in the network, and exercises for restoring content from nodes in the network to a server at a member institution.

The most rigorous and therefore most useful tests of the networks have not been planned exercises, however, but actual events of the type that the networks are designed to withstand. The most serious system failure occurred in the MetaArchive Cooperative in late November 2007, when the server room at the Woodruff Library at Emory University experienced a series of power failures that affected both of the uninterruptible power supplies that serviced the primary node of the MetaArchive Cooperative's LOCKSS network and the AX100 storage array that was attached to that node. This array stored all of the data for the archival units managed by the server. Working over several days, Emory University IT staff attempted to restore the primary LOCKSS server and the storage array in the face of successive power failures that undid the repairs they had performed and caused irreparable damage to the network filesystems. They eventually succeeded in repairing the primary server and the storage array by replacing damaged hardware and replacing the damaged filesystems with filesystems from other nodes in the MetaArchive network. LOCKSS is designed explicitly to enable this type of recovery.

The repair process demonstrated the viability and elegance of LOCKSS-based preservation networks. Restoring the Emory-based node took over a week and resulted in some hard-won lessons about restoring a LOCKSS node in the face of multiple and successive system failures. The chief lesson had to do with repairing the damaged filesystems. The corruption of filesystems housing archival units or the failure of the storage media containing those filesystems is among the most likely recoverable failures of any LOCKSS node. Although Emory's attempts to recover the filesystem were delayed by a second power failure, the process consumed a great deal of time, and the filesystems were likely beyond recovery even before the second failure occurred. Checking and repairing three large filesystems on AX100 hardware took far more time than the eventual solution of simply reformatting the disks and re-crawling the data from remote sources. It is obvious that after severe catastrophes in which hardware or hosting facilities are wholly destroyed, the best course of action is to recover archival units from remote sources. This is in fact a fundamental design principle of LOCKSS. An unanticipated lesson of these incidents, however, was that even less serious hardware or filesystem failures are sometimes more easily repaired using remote sources. Although the LOCKSS system provides for the automatic detection and repair of corrupted archival units, the MetaArchive Cooperative recommends recovering damaged filesystems from other machines in a network rather than by attempting filesystem repairs, as the former method is likely to result in less node downtime.

ADPNet has also experienced and dealt with unplanned system problems. The network's ability to restore content to one of its members was tested in the spring and summer of 2008, when several archival units that had been staged for harvesting into the network on a Web server at the University of Alabama were lost when that server was inadvertently decommissioned. The job of extracting the content in question from the University of Alabama's ADPNet cache and restoring it to another Web server at Alabama was originally assigned a low priority by the LOCKSS support team because the content had been ingested into the ADPNet network, but it soon became clear that this was an excellent opportunity to test getting content out of the network and then re-harvesting it back into the network under real-life conditions. It took some weeks of experimenting with and debugging a custom script, but the three archival units were successfully restored from the network in their original state and at their original location. In the course of two weeks (the interval that LOCKSS caches try getting the manifest page from archival units that are not supposed to be permanently offline), each LOCKSS cache in the ADPNet network fetched the manifest page for each of the units, realized that they were once again available for harvesting, and successfully harvested them back into the preservation network.

Precisely because they were unplanned and unanticipated, these and other incidents demonstrated the resiliency of LOCKSS-based preservation networks in the face of serious hardware and software failures. That is, the type of failures that can be counted on to occur eventually in computer-based systems, no matter how well-designed they may be. They therefore demonstrated, in the most convincing fashion, the value of the strength-in-numbers principle as it is applied to digital preservation and embodied in DDP networks.

DDP and Private LOCKSS Networks: Practical Issues

The experience of the MetaArchive Cooperative and ADPNet suggests that LOCKSS-based DDP networks are an elegant and affordable way to preserve locally created digital content, regardless of the type of institution or the nature of the content to be preserved. If a group of institutions in one of the poorest states in the United States can set up and sustain a robust digital preservation network, then presumably other institutions in other states and countries can do it too.

This raises a practical question: How does a group of institutions go about setting up a LOCKSS-based preservation network? A good first step would be to download and read a copy of *A Guide to Distributed Digital Preservation*, the MetaArchive Cooperative's first book—it was published in 2010 by the Educopia Institute, and it is the first comprehensive guide to the subject. The *Guide* is available for free as a PDF file from the MetaArchive Web site.[14]

The first requirement for a PLN is a quorum of at least six institutions that have locally created digital content they would like to preserve and that have agreed to work together to create the network and to allocate sufficient resources to sustain it over the long term. A PLN may have more than six members—MetaArchive, ADPNet, and most other PLNs do—but six is the recommended minimum to ensure network robustness in the event that one or two nodes experience a simultaneous failure.

The second requirement is a policy or governance document. This document contains the rules for running the network and spells out the rights and responsibilities of the network members. When the MetaArchive Cooperative began its work in 2004-2005, there were no governance documents for collaborative digital preservation networks to use as models, so the members had to draft their own from scratch, with some help from legal counsel at one of the member institutions (Dwayne Buttler, University of Louisville) and *pro bono* contributions from a private law firm in Atlanta. Thanks to MetaArchive's work and work by other preservation initiatives in North America, there are now at least three publicly available governance documents that nascent preservation networks can copy or adapt to their purposes: the MetaArchive Cooperative Charter, the ADPNet Governance Policy, and the COPPUL PLN Governance Policy. All of these documents are publicly available on the Web sites of the three PLNs.[15] Other collaboratives are encouraged to use them as models.

Finally, setting up a distributed digital preservation network requires money, either in kind or in cash. Distributed digital preservation is far less expensive than re-creating damaged or destroyed collections, but it is not without cost. In general, the costs can be divided into four categories: hardware, staff time, communication, and membership fees.

Hardware first. Every preservation site in a PLN needs a dedicated LOCKSS server computer, or LOCKSS cache. LOCKSS will run on inexpensive, even surplus or superannuated equipment, but we have found that it runs best on up-to-date servers with at least several terabytes of expandable storage capacity. These can be physical servers or virtual environments, including VMWare and cloud-based options. Although prices are falling, physical servers with the appropriate capacity typically cost between USD$2,000-USD$5,000, depending upon the vendor.[16] Remember too that as a digital preservation network grows, additional storage space needs to be purchased and that hardware must be refreshed at regular intervals.

Staff time is needed is manage the LOCKSS equipment and to write the documentation and instruction sets (manifest pages and plugins) that LOCKSS uses to identify available content and harvest it into the network. The total commitment in staff time is not very large—typically several hours per month or even less—but it is an expense and needs to be considered at the outset. Communication costs are negligible, at least in our experience. The MetaArchive Cooperative conducts weekly conference calls and holds an annual meeting of the Cooperative's Steering Committee. ADPNet conducts monthly conference calls and holds an annual meeting of the network's Steering Committee. COPPUL conducts "mostly monthly" Skype calls. All three networks have listservs, and most routine business is conducted by e-mail.

This brings us to membership fees, the single most expensive item on the list. There are two types of membership fees in PLNs: the annual LOCKSS Alliance fee, which is usually required but may be waived at the discretion of the LOCKSS administration, and network membership fees, which are optional. The LOCKSS Alliance fee is based on the Carnegie Classification system for colleges and universities in the United States and currently ranges from USD$1,080 per year for small, two-year institutions to USD$10,800 per year for large research universities. Obviously, this is a substantial expense, and it has put LOCKSS-based digital preservation beyond the reach of smaller, poorly-resourced institutions—that is, precisely those institutions whose digital collections are most vulnerable to loss.

In an attempt to eliminate this obstacle to membership, the Alabama network worked out an agreement with LOCKSS that will permit institutions to join the network for a graduated annual membership fee without also having to join the LOCKSS Alliance, as long as the network delivers an previously agreed-upon amount for the year to LOCKSS to pay for continued software development and technical support. The new ADPNet membership system consists of four membership categories with progressive annual membership fees, base storage allocations in the network and fees for increasing that allocation, different levels of technical and administrative responsibility, and different levels of representation on the ADPNet governance bodies. Specifically, the four ADPNet membership categories are: Anchor (base annual membership fee: USD$5,000; base local data allotment: 1.5TB); Host (base annual membership fee: USD$2,400; base local data allotment: 500GB); Participant (Large) (base annual membership fee: USD$800; base local data allotment: 1.5GB); and Participant (Small) (base annual membership fee: USD$300; base local data allotment: 500MB).[17]

The new four-tiered ADPNet membership system was designed to address three issues. First, by divorcing membership in ADPNet from membership in the LOCKSS Alliance, it was designed to make participation in the network possible for smaller, poorly resourced institutions that cannot afford the LOCKSS Alliance membership fees. Second, it was designed to enforce the principle of "use more, pay more" by making membership fees commensurate with usage of the network. Third, and in that connection, it was designed to address the "free rider" problem that was identified by the authors of the BRTF-SDPA Final Report and which they defined this way:

> free-rider problem: a situation arising when goods are nonrival in consumption, when benefits accrue to those who don't pay for them. For example, the costs of preserving digital assets may be borne by one organization, but the benefits accrue to many.[18]

The new ADPNet membership system ensures that all the members pay something in order to belong to the network. At the same time, the less-expensive membership categories were designed to persuade institutions that might otherwise opt out to participate. Evidence to date suggests that the system is working as intended. Two public libraries—the Birmingham Public Library and the Huntsville-Madison County Public Library—joined the network at the end of 2011, the first at the Host level, the second at the Participant (Small) level. The network now consists of a state agency, five large or medium-sized research universities, a small liberal-arts college, and two public libraries—a fairly diverse membership. This early evidence suggests that the system of graduated membership fees will be successful; we hope that it can serve as a model for other digital preservation networks that are facing the same problem.

The MetaArchive Cooperative has grappled with some of the same issues, but has created a substantially different organizational model. From its founding, the Cooperative has sought to provide a distributed digital preservation environment that does not depend upon any particular technology and that is both administratively and technically separate from other initiatives. MetaArchive and its members have long held that the digital preservation field is in its infancy, particularly in terms of its technological development. Rather than defining its aims by a technical solution, the Cooperative has focused primarily upon its mission—to enable memory

organizations to own and control the process of preservation by embedding the necessary knowledge and infrastructure locally. From a business continuity perspective, the Cooperative determined in 2004 that it needed to maintain a technical environment and support apparatus that was physically and administratively separated from other initiatives, including the LOCKSS team. The Cooperative hosts a small central staff that provides program management, plugin development assistance, software development, and systems administration. Unlike ADPNet, COPPUL, and most other PLNs,[19] the Cooperative maintains its own LOCKSS network and pursues a considerable amount of research and development work therein.

This difference is reflected in the way that the Cooperative handles its membership fees. The Cooperative encourages all members to join the LOCKSS Alliance, but currently, non-research level institutions (in Carnegie terms) are not required to join the LOCKSS Alliance in order to participate in the MetaArchive network. MetaArchive charges an annual membership fee to all of its participants using a three-tiered approach. The "Sustaining Member" category, the highest level of membership, includes a position on the Steering Committee and is USD$5,500 per year. The second level of membership is the "Preservation Member" category, and is USD$3,000 per year. The "Collaborative Member" category, which enables consortia of institutions that already host their content in a collaborative repository to join the network as a unit, is USD$2,500 plus a nominal per-participant fee (usually $100 or less).[20] Membership fees are used to support the Cooperative's administrative, collaborative, and software-development activities, which are more substantial than those of most other PLNs. This structure serves to protect the Cooperative from overdependence upon the LOCKSS team and helps to ensure the Cooperative's business continuity over the long term.

Continuity: Some Guiding Principles

The MetaArchive Cooperative and ADPNet have identified a number of principles that contribute to continuity and economic sustainability. Briefly, the main ones are as follows:

1. Whenever possible, use open-source solutions (e.g. LOCKSS)—not necessarily because they cost less than commercial solutions, although generally they do, but because they can be managed and modified locally. This is an important consideration if one believes that cultural heritage organizations should retain control of and access to the digital content they want to preserve while minimizing their dependence on third-party solutions.

2. Whenever possible, take advantage of existing administrative infrastructure. There is a corollary here: whenever possible, avoid creating new administrative infrastructure. As was mentioned above, ADPNet is part of the Network of Alabama Academic Libraries (NAAL), an existing state agency. For various reasons, the MetaArchive Cooperative decided to create a new administrative entity (the Educopia Institute in Atlanta, Georgia) to manage that network, but that decision was necessitated by the network's geographic dispersion across a number of states and the absence of a satisfactory existing administrative home. In the event, this arrangement does not seem to have impeded the network's growth. On the contrary, basing the administration of the network with a neutral agency seems to have allayed concerns about institutional favoritism (and fluctuations in institutional commitment) and increased the network's attractiveness to potential members.

3. Aim for a lightweight administrative structure. Like any other form of administration, administering a digital preservation network costs time and money, and it is therefore advisable to keep the administrative structure as simple as possible. ADPNet and the COPPUL PLN each have just two committees: a steering committee for policy questions and a technical committee for hardware and software issues. The MetaArchive Cooperative has a similar administrative structure. The networks have different communication schedules: due to its size and relative complexity, MetaArchive holds weekly conference calls, the COPPUL PLN meets via Skype every other week, and ADPNet has monthly conference calls. A lot of business in all three networks is conducted by e-mail. The idea is to make digital preservation a routine, low-maintenance, and integral part of an institution's information-management activities.

4. Delegate as much decision-making power as possible to the individual member institutions. They know their digital collections best, and are best able to set preservation priorities.

5. Broaden "ownership" of the network by involving all the network members in management and administration. The chair of the ADPNet Steering Committee—the network's policy-making body—rotates among the participating institutions every year or two. This helps to ensure a flow of fresh ideas and approaches and gives all of the members a stake in the network's success. The same arrangement obtains in the COPPUL PLN. Management of the MetaArchive Cooperative tends to be concentrated in the central office that was created for that purpose, but the member institutions are represented on the network's steering committee.

6. Finally, a perhaps-controversial and counterintuitive principle: resist spending a lot of time working on "business models" or devising detailed financial justifications for digital preservation. Such activities may be necessary at the national level or for very large and complex organizations (e.g. national libraries and archives), but they are less useful at the local level. The very fact that institutions have invested substantial resources in creating digital collections and have a professional and fiduciary interest in protecting that investment by preserving those collections is reason enough to institute a digital preservation program. Doing so will require planning and the apportionment of responsibilities, but it should not require elaborate and time-consuming justifications. If it does, that itself may be a sign that long-term institutional commitment is lacking.

Conclusion: Continuity and Sustainability

The MetaArchive Cooperative, ADPNet, and other Private LOCKSS Networks have shown that caching identical copies of digital content in multiple, geographically dispersed locations increases the survivability of that content and ensures continuity of digital library services. In planned and unplanned disaster-recovery exercises, the networks have demonstrated the solidity of the principles on which Distributed Digital Preservation is based. This is a noteworthy achievement in its own right, but there are other benefits as well. Robert Fox of the University of Notre Dame has identified a number of "key advantages" of peer-to-peer digital preservation networks, including "garner[ing] support from like-minded institutions and rais[ing] the awareness level regarding the preservation of key digital assets"; "the potential to increase the knowledge base required to maintain the preservation systems being used"; and "increas[ing]

the opportunity for validity checking, especially in systems that use 'voting' as a mechanism for checking file integrity".[21] Digital preservation networks also offer excellent opportunities for international collaboration—a point that seems especially worth emphasizing at an international conference of librarians from many countries. Geographic separation of LOCKSS nodes is one of the core features of DDP, and the more far-flung the LOCKSS caches are, the more survivable the network will be. It is hoped that the points raised in this paper will help to persuade other institutions that distributed digital preservation is a technologically viable option for their digital collections.

And an affordable one. Digital preservation is widely perceived to be a complex and expensive undertaking, requiring years of planning and large infusions of money and other resources. As Fox put it, the issues surrounding long-term digital preservation "are daunting not only owing to the complexity of the topic, but also the time commitment that would be required to implement very robust preservation systems".[22] This perception may be true in some cases, but it need not be. The experience of the DDP networks in North America suggests that it is possible to build robust, scalable, and economically sustainable preservation solutions with relatively modest resources. Moreover, it is possible to extend this solution across different kinds of institutions in different states, provinces, and countries. The MetaArchive Cooperative is a truly international preservation network, with institutional members in Brazil, Spain, and the United Kingdom. The ADPNet-COPPUL relationship is an example of two self-sustaining DDP networks that are collaborating fruitfully across national borders. Finally, MetaArchive and ADPNet have demonstrated that LOCKSS-based DDP is capable of ensuring the continued availability of digital content even in cases of serious system failure. Taken together, these initiatives represent working examples of technologically resilient solutions and offer proof that it is possible to create affordable and sustainable preservation networks internationally.

Acknowledgments

The authors would like to acknowledge the useful comments provided by Dr. Katherine Skinner, the executive director of the Educopia Institute, on an early draft of this paper. They take full responsibility for the paper and its conclusions, however.

Sources

Alabama Digital Preservation Network (ADPNet): http://adpn.org/

AlabamaMosaic: http://www.alabamamosaic.org/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (BRTF-SDPA 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*: http://brtf.sdsc.edu/biblio/BRTF_Interim Report.pdf

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (BRTF-SDPA 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*: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Chronopolis: http://chronopolis.sdsc.edu/

Council of Prairie and Pacific University Libraries (COPPUL) Private LOCKSS Network: http://www.coppul.ca/pln.html

Digital Curation Centre (DCC): http://www.dcc.ac.uk/

The Economist (2012). "History Flushed: the digital age promised vast libraries, but they remain incomplete". *The Economist* (April 28, 2012): http://www.economist.com/node/21553410

Fox, R. (2011). "Forensics of digital librarianship". *OCLC Systems & Services* 27 (4): pp. 264-271. http://dx.doi.org/10.1108/10650751111182560

Joint Information Systems Committee (JISC): http://www.jisc.ac.uk/

Lavoie, B., and & Dempsey, L. (2004). "Thirteen ways of looking at: digital preservation". *D-Lib Magazine* 10 (7/8): http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

LOCKSS: http://www.lockss.org/

MetaArchive Cooperative: http://www.metaarchive.org/

National Digital Information Infrastructure and Preservation Program (NDIIPP): http://www.digitalpreservation.gov/

National Digital Stewardship Alliance (NDSA): http://www.digitalpreservation.gov/ndsa/

Network of Alabama Academic Libraries (NAAL): http://www.ache.state.al.us/NAAL/

Skinner, K., & Schultz, M., eds. (2010). *A Guide to Distributed Digital Preservation*. Atlanta, GA: Educopia Institute: http://www.metaarchive.org/GDDP

References

[1] Fred Kaplan, *The Wizards of Armageddon* (New York: Simon and Schuster, 1983).

[2] "History Flushed: the digital age promised vast libraries, but they remain incomplete", *The Economist* (April 28, 2012): http://www.economist.com/node/21553410

[3] Library of Congress, Digital Preservation: http://www.digitalpreservation.gov/; National Digital Stewardship Alliance (NDSA): http://www.digitalpreservation.gov/ndsa/

[4] Digital Curation Centre (DCC): http://www.dcc.ac.uk/

[5] BRTF-SDPA Final Report, p. 9.

[6] Katherine Skinner and Monika Mevenkamp, "Chapter 2: DDP Architecture", in Skinner, K. and Schultz, M. (eds.), *A Guide to Distributed Digital Preservation* (Atlanta, GA: Educopia Institute, 2010): pp. 12-13.

[7] LOCKSS Program: http://www.lockss.org/

[8] Private LOCKSS Networks listing: http://www.lockss.org/lockss/Private_LOCKSS_Networks

[9] MetaArchive Cooperative: http://www.metaarchive.org/; Alabama Digital Preservation Network (ADPNet): http://adpn.org; COPPUL Private LOCKSS Network: http://coppullockssgroup.pbworks.com/w/page/11478105/FrontPage

[10] Network of Alabama Academic Libraries (NAAL): http://www.ache.state.al.us/NAAL/

[11] For more about the 2007 tornado in Enterprise, Alabama, see: http://en.wikipedia.org/wiki/Enterprise,_Alabama

[12] For more about the "2011 Super Outbreak", see http://en.wikipedia.org/wiki/April_25-28,_2011_tornado_outbreak

[13] Documenting the American South: http://docsouth.unc.edu/; Digital Library of Georgia: http://dlg.galileo.usg.edu/?Welcome; AlabamaMosaic: http://www.alabamamosaic.org/

[14] The *Guide* is available at http://www.metaarchive.org/GDDP

[15] These governance policies are available at the following locations: http://adpn.org/resources.html; http://coppullockssgroup.pbworks.com/w/page/11478105/FrontPage#GovernancePolicy; http://www.metaarchive.org/documentation

[16] See the MetaArchive Cooperative's technical specifications document at http://www.metaarchive.org/documentation

[17] For details on the different levels of membership, see the "ADPNet Membership Model" at http://adpn.org/resources.html

[18] BRTF-SDPA Final Report, p. 107.

[19] Notable exceptions are Data-PASS (ICPSR/Odom Institute: http://www.data-pass.org/) and LuKII (Germany: http://www.lukii.hu-berlin.de/), both of which host their own infrastructures and undertake substantial research and development activities.

[20] See: http://www.metaarchive.org/how-to-join. Starting in 2012, the membership fee for the Collaborative category is calculated on a case-by-case basis in accordance with the number of member institutions in each consortium: see http://www.metaarchive.org/costs/

[21] Robert Fox (2011). "Forensics of digital librarianship", *OCLC Systems & Services* 27 (4): p. 268.

[22] *Ibid*., p. 271.