

CARLI and Digital Preservation: A White Paper

1. Introduction

In 2009, while preparing a jointly sponsored forum on digital preservation, the chairs of the Digital Collections User Group and the Preservation Working Group noted that there is no CARLI plan for the long-term preservation of, and access to, member- and consortially-developed and funded shared digital resources. Members of the Digital Collection Users' Group and the Preservation Working Group formed a joint subcommittee in 2009 to investigate CARLI member libraries' digital preservation needs. As an informal group, the subcommittee's investigations were limited and did not include inquiring about pricing models, for example.

Now that national and international standards for building scalable systems for digital preservation are available and in use, the subcommittee believes CARLI has the opportunity to take a leadership role for its member institutions and investigate collaborative opportunities and solutions that provide members with the resources, options, and models that libraries throughout the consortium might use to preserve their digital resources. CARLI can leverage knowledge drawn from the current work of other libraries and consortia in a cost-effective manner. Encouraging, facilitating, and providing logistical support to libraries working on a collaborative digital preservation project is well within CARLI's strategic priorities.

All CARLI libraries share the challenge of digital preservation, and we ask the CARLI Board to investigate collaborative opportunities and solutions to answer this challenge.

2. Executive Summary

This white paper recommends that the CARLI Board formally establish a Digital Preservation Task Force, with an initial three-year term that may be extended if necessary, to provide leadership to CARLI for building a digital preservation program that will appropriately serve CARLI member libraries.

The two key charges of this task force would be:

1. **Create a series of tiered educational programs ranging from basic to advanced for CARLI member libraries, in preparation for the establishment of a CARLI-wide digital preservation program.** The Preservation Working Group and the Digital Collections Users' Group should be instrumental in presenting educational content, and should work with CARLI to support the bringing of outside experts to Illinois or online as part of this educational mission.
2. **Investigate options for consortial-level digital preservation activities, such as a dark archive.** This investigation should include pricing models at all tiers of a digital preservation program implementation and the necessary CARLI support required for such activities. This investigation should lead to presenting a formal set of recommendations to the CARLI Board for establishing and funding a digital preservation program for participating CARLI libraries.

3. Subcommittee Membership

- Lynne Thomas, subcommittee chair (Northern Illinois University, chair of the Preservation Working Group)
- Ellen Corrigan (Eastern Illinois University, chair of Digital Collections Users' Group)
- Kevin Ford (Columbia College, representative from Digital Collections Users' Group)
- Jennifer Hain Teper (University of Illinois at Urbana-Champaign, representative from the Preservation Working Group)
- Julie Mosbo (Southern Illinois University Carbondale, representative from the Preservation Working Group)
- Laurie Sauer (Knox College, representative from Digital Collections Users' Group)
- Nathaniel Wilson (National-Louis University, representative from Digital Collections Users' Group)

CARLI staff liaisons

- Elizabeth Clarage
- Jessica Gibson
- Amy Maroso

4. What is Digital Preservation?

Digital preservation is a commitment to maintain long-term access to digital objects through standardization, migration, and replication of those objects on numerous servers in multiple locations. This is true regardless of whether the objects were converted from analog sources or born digital. The definition of “long-term” varies widely but can be anywhere from five to fifty or more years. Simply put, a backup tape is not a digital preservation plan.

These kinds of preservation efforts differ from traditional preservation techniques in that they must be proactive and continuous. Proper storage and “benign neglect” (i.e. minimal handling) is usually enough to maintain physical media-based information on paper or polyester film for decades, if not centuries. However, preservation techniques that are appropriate for traditional sources will not adequately protect digital materials. All digital information is vulnerable to deterioration and loss in a process called “bit rot” because of the inherent vice of digital storage materials like hard drives, CDs, DVDs, thumb drives, etc. Factor in software and hardware changes that threaten to make our information unreadable and we have a recipe for significant data loss if no action is taken.

Preserving these digital objects for the long term is a complicated process, but there is a shared set of ISO standards, and an intellectual model: the OAIS Model for digital preservation. The OAIS model describes the process of caring for data through its entire lifecycle in a digital preservation system: bringing it into the system, standardizing it, checking it for errors, assigning metadata, preserving multiple copies of it that check with each other to verify their authenticity and completeness, and managing and retrieving those objects as necessary.

Digital objects have become an essential part of our historical and cultural records and the growth rate of digital collections will only increase over time. Over the past 10 years, CARLI member libraries have collected and created a significant number of digital objects through

digitization projects and purchases. These resources have been tremendously beneficial to users and have become an essential resource for many libraries.

Despite their value and potential, digital records are some of the most fragile materials in collections. Bit rot is only one example of how digital materials can be corrupted or lost forever. Many digital records are the target of malicious attacks, are poorly copied or stored on media that is no longer readable by current hardware and software, or are simply deleted because data managers do not see them as important. As but one example, presidential scholars cannot study the impact of the Internet on the 1996 presidential election—the first conducted during the Internet Age. The record is incomplete since the web-based digital documents created during that election are, according to the Library of Congress, now “lost to history;” the data disappeared before the Internet Archive, established that same year, could preserve it. In the absence of a viable digital preservation plan, all digital records are in danger of the same fate as those lost digital records, both through bit rot and through disasters such as fires or floods. We must keep investing in infrastructure to move these materials forward and maintain their history.

As a consortium, CARLI has a tremendous opportunity to help protect the digital records of its member libraries by supporting elements of digital preservation that are beyond the reach of many individual institutions. Numerous speakers at the December 2008 Northeast Document Conservation Center (NEDCC) “Persistence of Memory: Sustaining Digital Collections,” held in Chicago, emphasized that both the necessary human and financial resources for digital preservation are too great for most libraries to manage individually; digital preservation works best and is most cost-effective when libraries and campuses collaborate with one another.

Nancy McGovern, an internationally known expert on digital preservation who serves as the digital preservation officer for the Inter-University Consortium for Political and Social Research (ICPSR), the organization now hosting the Cornell Workshop on Digital Preservation, describes digital preservation as a three legged stool, with each leg—technology, organization, and resources (both human and financial)—supporting the entire program equally. CARLI's history of fostering collaboration on projects, such as the statewide online catalog, point to an ability to leverage all three legs of the stool. Many institutions cannot build up and maintain all three legs by themselves and CARLI is in a perfect position to support elements of digital preservation programs throughout the state of Illinois. CARLI can also act as a model for other consortiums throughout the country.

5. Why Should CARLI take on Digital Preservation?

5.1 Digital preservation aligns with the CARLI mission

In 2006, the CARLI Image Database Task Force noted in their final report that “the CARLI consortium ought to consider what, if any, role it will play in long-term preservation of content that its members make available through the consortium's portal and search/discovery mechanisms.” Since that report CARLI has moved forward with numerous digitization projects through licensing CONTENTdm for members, its partnership with the Open Content Alliance, and encouraging member libraries to individually ensure long-term access to their subscribed resources through services such as Portico.

That CARLI encourages academic libraries to acquire, create, and share digital objects in the furtherance of our shared teaching, learning, and research goals is commendable. However, up until now there has been no concurrent commitment to the long-term preservation of and access to these member- and consortially- developed and funded shared digital resources. Just as CARLI has encouraged and supported its libraries with the development of digital objects, it is time for CARLI to take the next step and lead the way in developing a digital preservation infrastructure to help libraries preserve their valuable digital assets.

5.2 Wide-scale collaboration on a digital preservation infrastructure is economically desirable

The alternative to wide-scale collaboration is for each institution to develop its own digital preservation infrastructure, which includes hardware, software, network infrastructure, and personnel with technical expertise. Given the amount of resources required for such an effort there are few, if any, CARLI institutions poised to be able to do this.

Even if it were possible for each CARLI institution to implement valid and reliable digital preservation infrastructures on their own, it makes more economic sense to eliminate redundancies by creating a consortium-based digital preservation program. Cooperation in this case will reduce redundancy; much like the centralized installation of the Voyager integrated library system (ILS) reduces the need for each CARLI member library to maintain its own ILS infrastructure. In an informal survey conducted by this subcommittee (see Appendix A), 14 out of 25 respondents said their libraries are interested in participating in a consortium-based digital preservation program, and another seven indicated they would possibly be interested in such an effort.

5.3 CARLI infrastructure and culture of cooperation already exist to facilitate implementation of a digital preservation infrastructure

CARLI already has the resources and communications systems in place to coordinate the necessary work for the creation of a digital preservation infrastructure, which would be administered through CARLI, and in which member libraries would work together to provide technology and human and financial resources. The CARLI committee structure could easily accommodate an additional body charged with responsibility and oversight for this effort.

Equally important is the emphasis CARLI places on developing a culture of cooperation. Collaboration and resource sharing are two of CARLI's important strategic principles. CARLI member libraries are accustomed to working consortially and have experience with collaborative purchases, creating and acquiring resources through the consortium, and sharing materials. Creating a consortium-based digital preservation infrastructure would work to maintain and enhance the culture of cooperation already in evidence in much that CARLI does.

6. Initial investigations

6.1 Informal survey of CARLI members

The initial meeting of the CARLI Digital Preservation subcommittee took place September 29, 2009. The meeting focused on discovering the digital preservation needs of the libraries within Illinois and the best way for CARLI to support those needs. Additional work was done through

follow-up telephone conferences during fall and winter of 2009. CARLI member libraries represent a variety of collection types and collection needs. To identify the educational and digital preservation needs of these collections, the subcommittee decided to develop a set of informal questions to ask a cross-section of the CARLI member libraries.

To understand the needs of all types of library comprising CARLI's membership, the subcommittee selected institutions from each of the Carnegie levels represented by the CARLI member libraries: research, masters, baccalaureate, associate, and special. A total of 31 institutions were selected. A questionnaire was sent to the Dean or Director of each library or to the appropriate individual in charge of managing digital content. Individuals were contacted either by phone or email according to the preference of the interviewer or interviewee. Respondents were made aware that their responses would be used collectively, with no institution being identified by their answers, and that the subcommittee would use the responses to create a report on digital preservation.

Please see Appendix A for the questions and results.

Key findings of the questionnaire:

6.1.1 Existing digital preservation programs at CARLI libraries

Only five libraries out of the 25 which responded said that their institution has some form of digital preservation program in place at the present time. The subcommittee was also interested in the broader question: If a library does not have a program, are there backup procedures in place for digital content? Sixteen libraries indicated that they do have backup procedures in place to safeguard their digital content.

Participants were asked if they had confidence in both their short-term and long-term backup procedures (Appendix A #6). Six libraries expressed a "good" level of confidence in their short-term backup procedures, while three libraries had "some" confidence; four libraries had little or no confidence in their short-term procedures. Confidence in long-term backup procedures was even lower: no library expressed having a "good" level of confidence, and only seven libraries expressed that they had "some" confidence in their long-term procedures, and for nine libraries their expressed level of confidence in their long-term backup procedures was low to non-existent.

6.1.2 Level of knowledge about digital preservation and digital preservation planning

One purpose of the questionnaire was to assess CARLI member institutions' current level of knowledge about digital preservation. When asked, "What do you know about digital preservation?" the 22 answers spanned a range of possibilities: five respondents reported knowing "not much" or "very little" about digital preservation (23%), 10 respondents reported having only general knowledge of principles and procedures or, to paraphrase, "enough to know what we're doing and where we mostly need to go, but no specific details" (45%), and only five respondents reported having a good understanding of digital preservation (23%). Of important note is the fact that there was a mix of institution types in all three categories, indicating that digital preservation is not an issue mastered by any one type of library. Only two libraries out of the 25 responses indicated they have a digital preservation plan (Appendix A #3). These two libraries are among the five libraries that reported they had a good understanding of digital

preservation. Four of the 25 libraries surveyed are currently developing plans and 16 have no plan at all.

6.1.3 Interest in a CARLI-based Digital Preservation Service

There was a generally positive response to the idea of a consortium-based digital preservation solution (“yes”=14; “possibly”=7, Appendix A #2)). To explore the idea of what kind of involvement a consortium might have in a digital preservation service (DPS), participants were asked, “Would a CARLI-based, state-wide system be more or less appealing to you than CARLI-sponsored participation in a national DPS?” (Appendix A #12). Eight respondents indicated that a CARLI-based system would be preferable, while another eight respondents had no preference or answered “it depends.” Only one respondent indicated a preference for a national system with CARLI sponsorship.

However, when asked about the form that a CARLI-based DPS might take, opinion was divided (Appendix A #9):

- dark archive only* (3)
- dark archive plus file migration services* (4)
- dark archive plus an institutional repository* (2)
- fully-functional institutional repository services (4)

But many respondents were also not as certain about a solution—six respondents were unclear or unsure what form a DPS should take. Further, 15 organizations did not respond as to whether a dark archive solution would be sufficient. Of the eight that did respond, five reported that a dark archive would be fine, two would require migration services along with a dark archive to participate in a consortial solution, and one is adamant about full institutional repository (IR) functionality.

6.2 Investigation of existing digital preservation programs

In addition to surveying CARLI member libraries, each member of the subcommittee selected a digital preservation program in the United States to research and contact. Those researched were the Alabama Digital Preservation Network (ADPNet), MetaArchive Cooperative, Dark Archive in the Sunshine State (DAITSS), Ex Libris Rosetta, and OCLC Digital Archive Service. These programs were selected because they have been proven to be stable, well-respected, and well-known in the field of digital preservation.

The ADPNet and DAITSS are examples of statewide consortium digital preservation programs. ADPNet is based on the LOCKSS (Lots of Copies Keep Stuff Safe) model. DAITSS developed their own open source software to serve as a dark archive platform. The two programs differ in the number and types of institutions that are members: ADPNet has eight members, and DAITSS consists of 10 universities and one college and does not allow membership for community colleges or private universities/colleges. Grant funding also is important in the development of

* See Glossary of Terms for definition

these consortial digital preservation system; both ADPNet and DAITSS systems are funded by Institute of Museum and Library Services grants.

The MetaArchive Cooperative, Ex Libris Rosetta, and OCLC Digital Archive Service are examples of services where individual institutions can apply to be members or pay a commercial vendor service. All three operate under the main idea of digital preservation: creating multiple copies and keeping them in multiple locations, but the MetaArchive Cooperative is the only system that states it is based on the LOCKSS model. The OCLC system appears to require the physical shipment of data and storage to a central location. Ex Libris Rosetta is based on the Open Archival Information System (OAIS) and is conforming to the trusted digital repository (TDR) requirements. The MetaArchive Cooperative currently has 15 members and is partially funded by a National Historical Publications and Records Committee grant. Information regarding who has or is currently using the Ex Libris Rosetta system or the OCLC Digital Archive Service could not be found.

Please see Appendix B for more information regarding the digital preservation services investigated by this task force.

7. Recommendation: Formation of Digital Preservation Task Force

The subcommittee recommends that the CARLI Board establish a Digital Preservation Task Force with a three-year term whose two charges are:

- Create tiered educational programs—ranging from basic to advanced topics—on aspects of digital preservation
- Investigate and evaluate options for consortial-level digital preservation activities, such as a dark archive, and deliver formal recommendation to CARLI board for consideration for implementation.

The immediate impact of this Task Force's formation will be through the education of the CARLI membership, which should encourage buy-in over the long term if a Digital Preservation Service is implemented.

The Digital Preservation Task Force would apply to the Program Planning Committee for program funding as necessary, and would need only nominal funding for task force meeting expenses (e.g., conference calls, lunches for in-person meetings).

7.1 Charge I: Educate CARLI libraries

Based on the results from the subcommittee's questionnaire, it is important to note that most libraries queried do not have digital preservation plans and only five of the 25 respondents have a good understanding of digital preservation. The questionnaire results reveal that many of the libraries already involved in even minimal digital preservation activities feel a significant lack of confidence in what expertise they do have. From these responses it becomes clear that education must be one of the first steps towards developing a comprehensive digital preservation program. The development of a digital preservation program without preliminary education would mean a significant challenge to CARLI to anticipate members' needs before members have fully understood them themselves. CARLI member institutions will benefit from formal education

and/or training in the fundamentals of digital preservation and in creating a digital preservation plan. **Therefore, this subcommittee recommends that CARLI support education and training for member libraries surrounding the issues of digital preservation.**

The educational needs of the membership are broad and must be met in a tiered manner. A need for instruction of both basic and advanced skills will be necessary. We propose that CARLI begin this area of instruction with low-cost, high impact introductory topics and then build to more advanced or in-depth topics. As this program is developed, CARLI should seek the expertise of known experts in the field (both local and national), for example Nancy McGovern (University of Michigan, ICPSR), the North East Document Conservation Center (coordinating agency for both the Persistence of Memory and Digital Directions workshops), Jerome McDonough (University of Illinois at Urbana-Champaign Graduate School of Library and Information Science), and Kevin Ford (formerly at Columbia College and now at Library of Congress). Additionally, there may be opportunities for partnerships in the state in the area of digital preservation, including a possible partnership with Nancy McGovern and the University of Illinois at Urbana-Champaign, as well as the development of training under the “Connecting to Collections” state-wide preservation program initiative.

Some suggested ideas for introductory topics that could utilize online education options such as webinars or asynchronous learning include:

- ICPSR Digital Preservation Tutorial: this is an existing and award winning web tutorial on the basics of digital preservation. Completion of this program could serve as a pre-requisite for other training opportunities provided by CARLI to ensure a basic level of understanding;
- Overview of open source tools available: this could present available digital preservation options, such as JHOVE (JSTOR/Harvard Object Validation Environment), that can be instituted and managed locally to meet a specific institution’s needs;
- An introduction for library administrators about the basics of digital preservation and how to begin conversations with library or campus IT to begin planning and implementation;
- How to write and implement an institutional digital preservation plan;
- Short online sessions covering such specific topics as:
 - How to regularize data formats;
 - An overview of metadata schemes;
 - File naming protocols;
 - The differences between an institutional repository and a digital preservation system.

More advanced training and educational sessions might address the following topics:

- How to select files (or file types) for preservation;
- Understanding data migration;
- Understanding checksums and data integrity;
- Exploring different methods of digital preservation, ranging from low cost options to more complex, higher cost options.

7.2 Charge II: Evaluate and recommend a specific DPS solution

Only two of the institutions that responded to the questionnaire have a digital preservation plan in place and only five institutions have implemented something resembling a DPS. More institutions have backup strategies and processes in place, but confidence in their back-up procedures for long-term digital preservation is low to non-existent. Furthermore, there was a great deal of positive responses from members surveyed to the idea of a consortium-based digital preservation solution, and, given that economies of scale for a DPS can be better achieved if such a service were built or managed at the consortial level, **the subcommittee recommends that CARLI sponsor work to evaluate and recommend a specific DPS solution.**

There are a variety of solutions available for digital preservation systems, such as DAITSS, LOCKSS, etc. There are also a number of consortial digital preservation models, such as MetaArchive, ADPNet, etc.

The proposed Digital Preservation Task Force will need to evaluate these solutions and models while considering the following factors:

- Who will manage a DPS adopted by CARLI? Will management responsibilities be centralized or shared by members of the CARLI network? What will the management structure of the DPS look like?
- How much will a DPS cost? Will this cost be shared by consortium members? If so, what factors will affect how much each member pays?
- Are there any additional ways of funding a DPS besides direct CARLI and member support?
- How difficult will it be for CARLI to implement a DPS?
- Who will manage the technological aspect of a DPS and what skills will be required for this work?
- What are the different features of DPS's available to CARLI?
- Which DPS will be able to handle the expected volume of material and traffic for this program?
- Which DPS system will be flexible enough to satisfy the current needs of CARLI members and expandable enough to satisfy future needs?
- How reliable are the DPS options available to CARLI?

One possible scenario might be that the Task Force recommends that CARLI initially plan for a dark archive DPS solution, with plans to introduce migration services and full-service institutional repository functionality at later stages.

8. Future Directions

It is our hope that the proposed Digital Preservation Task Force, at the end of its first three-year term, will have fostered an understanding of digital preservation at its most basic level across the CARLI membership and presented a viable, multi-step plan to implement a consortially-based digital preservation program.

Appendix A: Questionnaire sent to CARLI Members

Institutions contacted

The subcommittee organized the CARLI membership according to each libraries’ Carnegie classification—Research Institutions, Masters Institutions, Baccalaureate Institutions, Associate Institutions, Special Libraries. The subcommittee randomly selected 31 institutions to query based on the numbers of CARLI members in each Carnegie category. Five Associate and Research institutions were contacted; six Baccalaureate and Special institutions. Since the majority of CARLI libraries fall into the “Masters Institutions” category, nine institutions from that category were contacted.

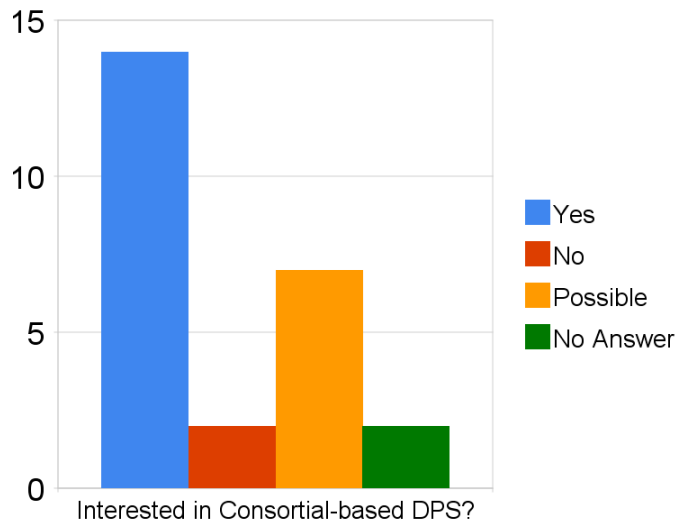
Of the 31 institutions contacted, 25 responses were received.

Questions and results

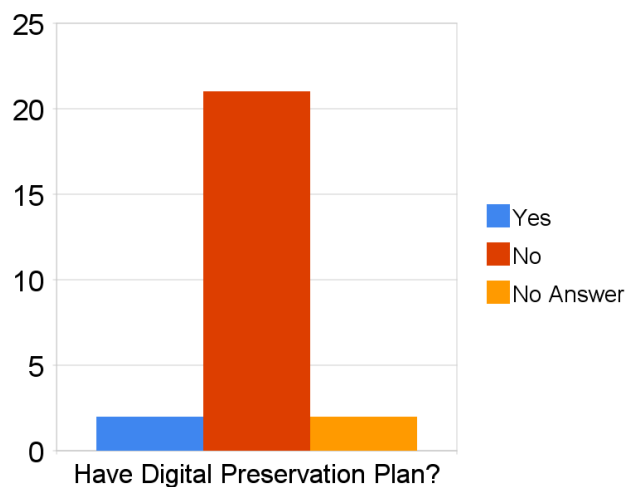
(1) What do you know about digital preservation? What is your current role in digital preservation in your library/on your campus?

Task force members asked this question primarily to ascertain if the respondent was the right person to answer the survey questions, and if not, to redirect the survey to the proper individual. Answers varied depending on the level of digitization being done at the institution.

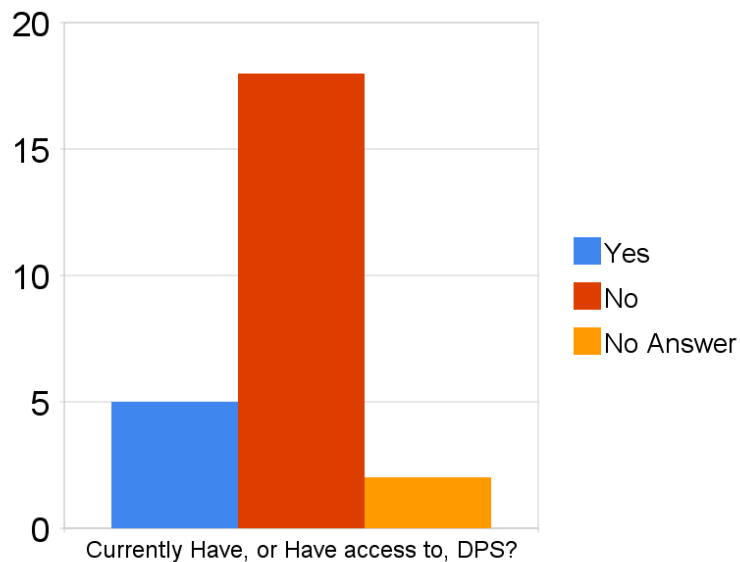
(2) Are you interested in participating in a consortial-based distributed digital preservation service (DPS)?



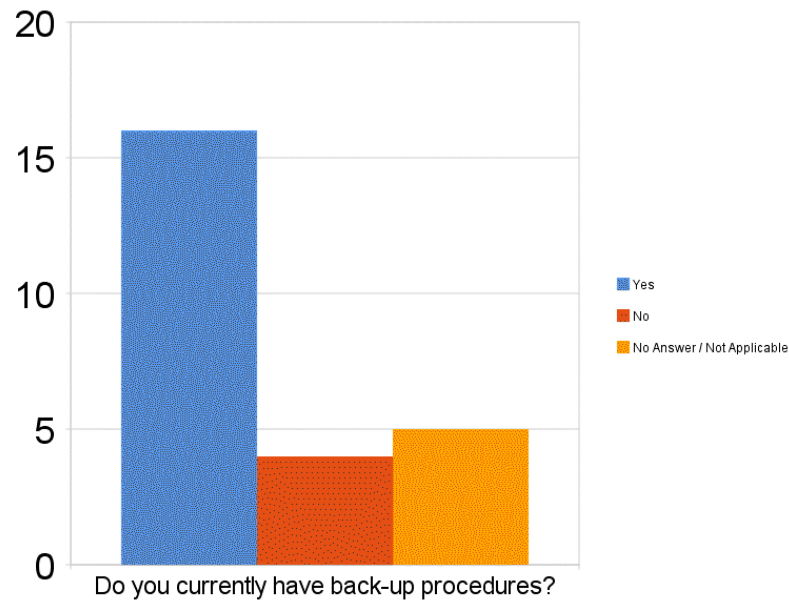
(3) Does your library or campus have a digital preservation plan?



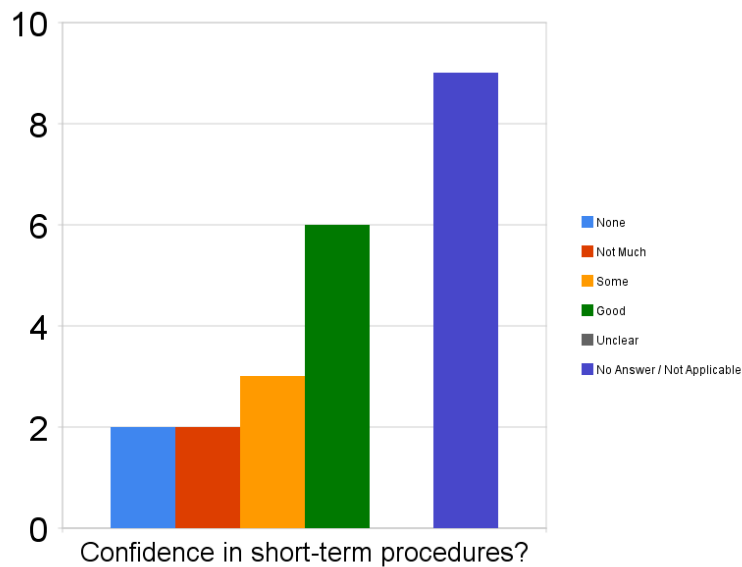
(4) Do you currently have a digital preservation service or access to a DPS? Follow-up: what system are you using? Examples: LOCKSS box, Portico, home-grown service, IR used as a DPS (i.e. DSpace, BePress, Fedora)?

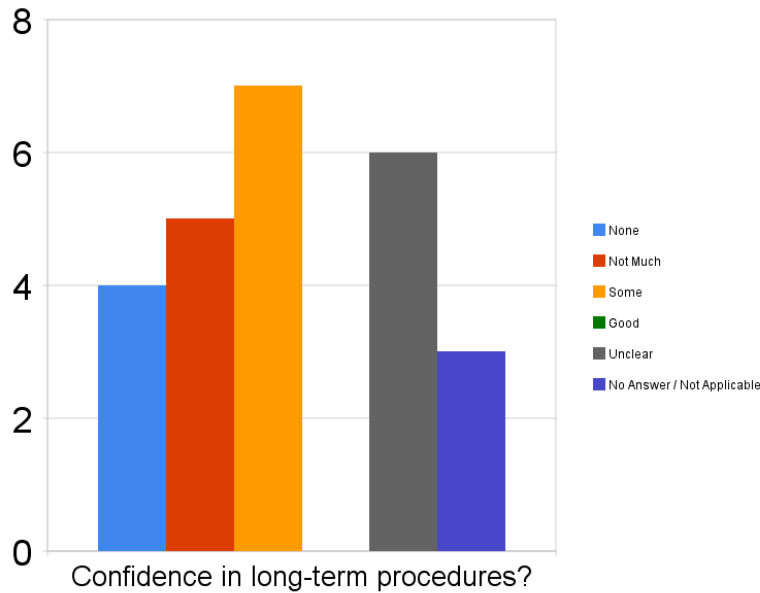


5) Do you currently have back-up procedures for your digital content? Explain what your current procedures are and are you doing this for all materials or just certain collections/files/file types/etc.?



6) How much confidence do you have in your current plan/procedures, both long-term and short-term?



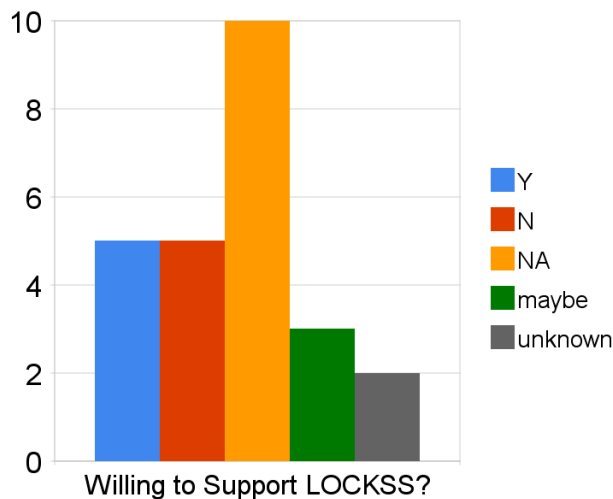


(7) How much material would you want to put in a consortial DPS? What types of material do you want to put in a DPS? (e.g. data sets, e-journals, ETDs, images, media files, text docs)

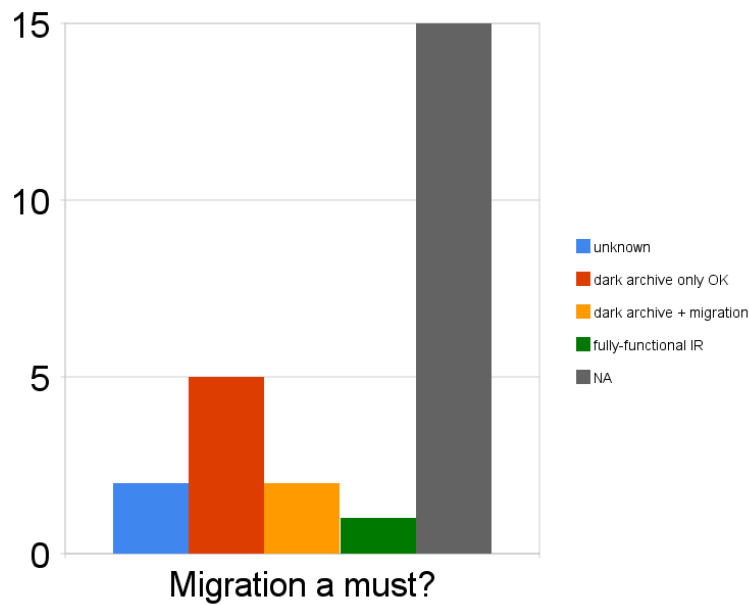
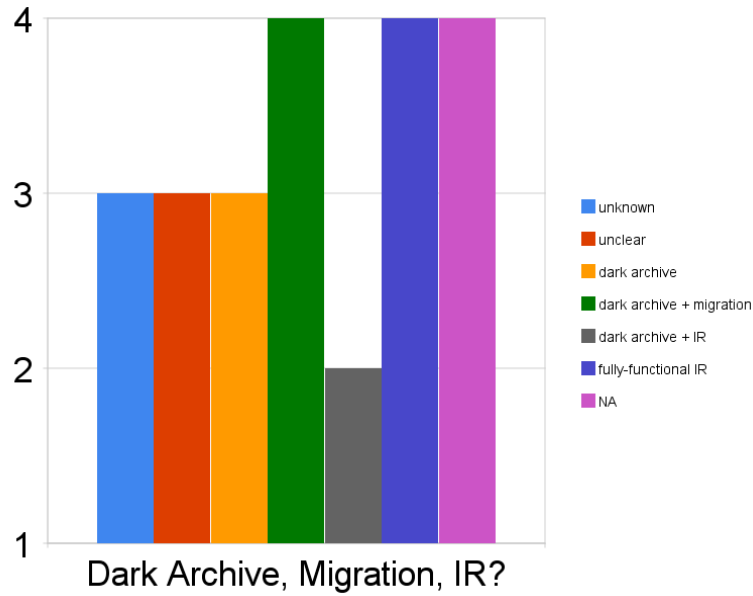
Answers to this question varied greatly:

- nothing (5 respondents)
- We don't have all that much content now
- No idea
- Depends on how much can control access to
- Limited quantities, for the first few years
- We have roughly 30K files
- A lot (3 respondents)
- At least 2.5 terabytes (TB)
- Twenty years of student records, administrative files, scholarly work, and data sets, e-journals, ETDs, images, media files, text docs
- At least 3 TB
- guessing 5-20 TB

(8) How much support do you get from the IT department on your campus for maintaining your DPS? Would they be willing to support a LOCKSS-based system?



(9) What types of DPS would you be most interested in? Dark archive only? Migration services? Full IR functionality (patron access to items within the system)? Would you be interested in a dark-archive-only service or would migration services be necessary for you to participate?



(10) What is your institution doing in terms of metadata (preservation, technical, administrative) for the items for which you are responsible?

For institutions with digital collections, answers to this question varied, but most respondents are doing little or nothing to collection preservation, technical and/or administrative metadata:

- Nothing yet (2 respondents)
- Currently – nothing. Would need guidelines for required metadata for a shared system.
- This is an area we have not spent much time on. I have briefly looked at several appraisal resources in the last two years.
- I'm trusting CARLI's and OCLC's servers.
- We do not add preservation, technical or administrative metadata to items, unless such metadata is automatically attached to a file during digitization.
- For the items ingested into the system, we collect descriptive, technical, and a limited amount of preservation/administrative metadata. For items not yet in the system, we only have a minimal amount of descriptive metadata about the objects.
- The [collection name deleted for anonymity] Collection only has a few metadata fields for this – file type, date digitized and file name.
- Primarily photo project in CONTENTdm. Other digital materials for library are MARC-based. Not in I-Share. OCLC, local ExLibris system for ILS.
- We are recording a minimal level of technical metadata for newly created digital objects as we begin our collections program. We're not addressing preservation issues in metadata at this time. Having CARLI-sponsored resources such as guidelines or resident consultative expertise on metadata would be useful to many colleges, in my opinion.
- For the internal photos, we do try and fully catalog the items with as much descriptive and technical information as possible. But again, this could be more robust in certain instances.
- Technical at time of digitization (CONTENTdm--looking for other possibilities); some preservation metadata (Secondary system or part of Dublin Core info).
- We create full preservation, technical and administrative data whenever possible. Most of it is created programmatically.

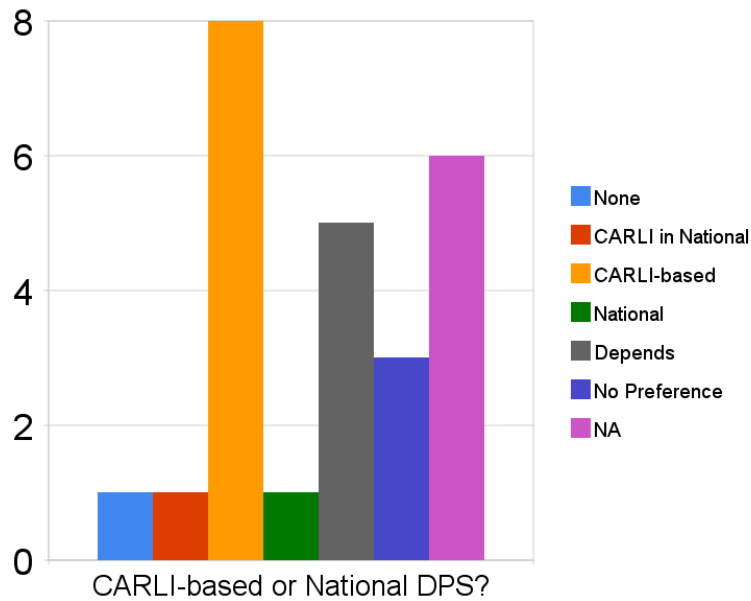
(11) How would you willing to be charged for this service—by space used or by size of institution?

The answer that garnered the most support was to charge by space used (six respondents). Three respondents favored charging fees based on the size of the institution. Other responses:

- If there were some added value services available we may be willing to pay.
- For us, pricing by size of institution would be in our favor, but I could see the need to based price on size of collection.
- Ala carte pricing would also be helpful. Since we are most likely not looking for patron access or even general staff access, these are services we wouldn't want to be paying for if they were offered.
- Whatever would be cheaper.
- If the service is limited to making use of one's own archive materials only, I'd say by "space used." If the service worked in a way that it provided access to the digital resources of other members, then by size of institution.

- My first instinct was to say by space used, since we don't have all that much digital content. But I know we will have more in the future. And we'll always be one of the smaller CARLI institutions. Maybe each participating school would be pay for x amount of space and then buy more? That wouldn't really be fair for larger institutions with lots of digital content though, would it? More discussion needed, I guess.

(12) Would a CARLI-based, state-wide system be more or less appealing to you than CARLI-sponsored participation in a national DPS?



Appendix B: Digital Preservation Systems Investigated

Our research did not extend to specific applications or programs used for managing a DPS (DSpace, Fedora, etc.)

Alabama Digital Preservation Network

Based on a private LOCKSS network (PLN), the Alabama Digital Preservation Network (ADPNet) was funded by a two year IMLS National Leadership Grant in 2006. The function of the system is to be a dark archive for locally created resources only. Member institutions must bring up/maintain a LOCKSS server, contribute digital content to the network and harvest from other network members, and commit to joining LOCKSS Alliance. The system has eight partnering institutions: Alabama Commission on Higher Education, Alabama Department of Archives and History, Auburn University, Spring Hill College, Troy University, University of Alabama, University of Alabama at Birmingham, and University of North Alabama.

Florida Digital Archive

Based on open source software, DAITSS (Dark Archive in the Sunshine State) is a preservation repository management application developed by Florida Center for Library Automation. The Florida Digital Archive (FDA) mission stated on their website is “to provide a cost-effective, long-term preservation repository for digital materials in support of teaching and learning, scholarship, and research in the state of Florida. In support of this mission, the FDA guarantees that all files deposited by agreement with its Affiliates remain available, unaltered, and readable from media. For supported formats, the FDA will maintain a usable version using the best format migration tools available.” The application provides full preservation includes bit-level preservation of originally submitted files and follows the Metadata Encoding and Transmission Standard model. Institutions that belong are called “FDA Affiliates.” DAITSS supports 10 universities and one college as part of their system. Only affiliates of the FDA are allowed to participate and to do so they must sign a formal agreement. Membership is not allowed for private universities and colleges or community colleges.

Orbis Cascade Alliance

The Orbis Cascade Alliance is a group of 36 institutions (universities, colleges, and community colleges) from Washington state and Oregon. In July 2009, Orbis Cascade Alliance with the Northwest Digital Archives produced a report entitled “Report and Recommendations of the Digital Program Working for Initial Elements of a Digital Services Program at the Orbis Cascade Alliance.” In this report, they discussed various options for implementing digital preservation services for the consortium. The options they investigated were: MetaArchive, OCLC Digital Archive, cloud storage*, and distributed offline storage. They rejected MetaArchive stating that the LOCKSS system used by MetaArchive was not scalable enough for their needs as it did not allow for storage of metadata with the associated digital files. In addition, running the system according to their specification would require a minimum of seven nodes which would be very expensive. The working group also rejected OCLC Digital Archive due to the cost and because it

* See Glossary of Terms for definition

requires physical media to be shipped. The cloud storage idea (example: Amazon's Simple Storage Service) had some merit because it is less expensive than MetaArchive and OCLC's service, but if the file integrity checking features were implemented, the costs would increase. They recommend keeping an eye on this market for possible use in the future. The report made a recommendation to adopt a low cost method they called "distributed offline storage." Libraries would put files they want to store on physical media and have multiple copies of those distributed to three other sites within the Alliance for safekeeping (in a powered-off state).

MetaArchive

Developed in 2004, the MetaArchive Cooperative's goal is to build a Trusted Digital Repositories using the open source LOCKSS software. The Cooperative includes libraries, archives, and various types of cultural heritage institutions who are interested in actively participating in the preservation of digital content rather than sending their content to an outside vendor for preservation. Institutions decide what collections to ingest into the LOCKSS systems and the files are then distributed to different locations geographically where the servers back up the materials and consistently check for file issues and needed repairs. MetaArchive has 15 participating institutions: Auburn University, Boston College, Clemson University, Emory University, Florida State University, Folger Shakespeare Library, Georgia Tech, Library of Congress, Pontificia Universidade Católica, Rice University, University of Hull, University of Louisville, University of North Texas, University of South Carolina, and Virginia Tech.

OCLC Digital Archive Service

The OCLC Digital Archive Service is a commercially managed digital storage. Provided in the OCLC system package is physical security where the system is monitored 24/7 by system operators, guards, and cameras; data security in which OCLC's security team reviews applications, systems, and procedures; data backup where copies of digital content is distributed geographically to separate and secure facilities; and disaster recovery. One of the key points that OCLC states regarding this system is the reported ease of integrating the system with workflows of CONTENTdm users. Their brochure states that the "master files are secured for ingest to the Archive using the CONTENTdm Project Client, the Connexion digital import capability and the Web Harvester." The system produces regular management reports (monthly) based on automated inspections and data verifications including: manifest verification, virus checking, fixity checks, and format verification.

Ex Libris Rosetta

Released in early 2009, Ex Libris Rosetta (previously known as the Digital Repository System) was developed in conjuncture with the National Library of New Zealand. The system is based on the Open Archival Information System (OAIS) and is conforming to the trusted digital repository (TDR) requirements; however, it is a commercial system. Ex Libris Rosetta has been built to manage digital collections of any size. The system can batch load and create multiple deposit hierarchies to manage the ingestion of new material. As stated in their brochure, the preserved files are delivered "via viewing applications for constantly evolving format types—supplied as part of Ex Libris Rosetta and third-party applications. Built-in integration capabilities enable the system to accept delivery requests from discovery and delivery applications such as the Ex Libris Primo® solution while ensuring that item-level access rights are enforced."

Glossary of terms

Bit Rot: The deterioration of digital files that occurs over time, when the media is not in consistent use. Data on electronic media will degrade, and eventually become unusable, when the media is not in use.

Cloud Storage: Saving of data to multiple, off-site, remote databases maintained by third parties. An Internet connection provides the connection to the off-site storage databases.

Dark Archive: A digital preservation service that maintains copies of digital objects directly accessible only by system administrators. A dark archive cannot be accessed by the public through Google or other means.

Digital Preservation Service: A system, or a group of separate systems working in concert, designed to provide continuing access to digital objects over a long period of time (envisioned as anywhere from five to 50 years). Examples include LOCKSS (a geographically distributed group of servers that work together) and digital preservation services provided by vendors like OCLC.

Institutional Repository (IR): An institutionally based system that showcases the intellectual work of a particular campus. IRs have a search and discovery component and are often the public access method of getting at materials that are also preserved in a dark archive.

Migration: Moving a digital object from one format to another in order to preserve it. This can range from moving a Word document to the newest version, to converting images to archival file types such as TIFF and JPG2000. The goal with migration is to maintain the formatting and character of the original object for the most part, but to add usability by bringing it up to current standards.