SURVEY RESULTS

EXECUTIVE SUMMARY

Introduction

The 2010 OCLC Research report, *Taking Our Pulse*, listed management of born-digital materials as the third biggest challenge facing libraries, special collections, and archives, after space and facilities. It has become a truism that the trickle of born-digital materials into special collections has become a flood. Increasingly, these materials do not have analog counterparts. Libraries and archives can no longer defer decisions about digital content to a later date. We must develop policies and procedures to operationalize the management of born-digital materials, or we risk losing the record of the recent past.

This survey sought to gather and promote emerging good practices for managing born-digital content and to highlight common challenges. The survey instrument focused in particular on staffing, ingest and processing workflows, storage procedures, and access and discovery methods. Sixty-four of the 126 ARL member libraries responded to the survey between February 22 and March 23 for a response rate of 51%. Fifty-nine of the respondents (92%) already collect born-digital content. The remaining five libraries are in the planning stages. The level of engagement with born-digital content was higher than anticipated by the survey team. An analysis of the responding libraries engaged with born-digital materials revealed they are larger institutions and therefore more likely to be pioneers in working with this content.

The management of born-digital materials is still relatively new for ARL libraries, and the survey results show that good practices and workflows are still evolving. New tools are emerging rapidly, and the once-solid line between digitized content and born-digital content is beginning to blur. Survey responses indicated that the library and archives profession lacks a common definition of what born-digital content is and a common understanding of who within the organization should manage this content.

Staffing and Organization

The survey asked how many library staff collect and manage born-digital materials, who has responsibility for storage-related activities, how staffing needs are addressed, and how staff gain the expertise required to manage these materials. No one staffing or organizational structure emerged from the survey responses, which again reflects the evolutionary status of born-digital management programs.

The number of staff working with born-digital archival content in the responding libraries ranges from less than one to 60 FTE. While archivists and librarians in institutional and government archives were the trailblazers in collecting this content, managing these materials now requires staff from digitization, digital curation, information technology, and institutional repository units. Respondents most frequently mentioned special collections/archives staff and library IT staff as having decision-making responsibility for selecting storage solutions, implementing and maintaining infrastructure, managing user authentication, estimating storage needs and monitoring usage, and budgeting. Many other units are also involved, including institutional IT, preservation, collections, administration, and consortia in a wide variety of combinations.

This organizational distribution may factor into how respondents have addressed staffing needs for

managing born-digital content. Almost all have used a combination of strategies, either adding that responsibility to existing positions (94%) or recasting an existing position (37%), and creating new positions (46%). Training strategies reflect the emphasis on retooling the skill sets of existing positions. Conferences, on-the-job training, workshops, and independent study are the primary methods staff use to develop their expertise with born-digital content.

Born-Digital Materials Collected

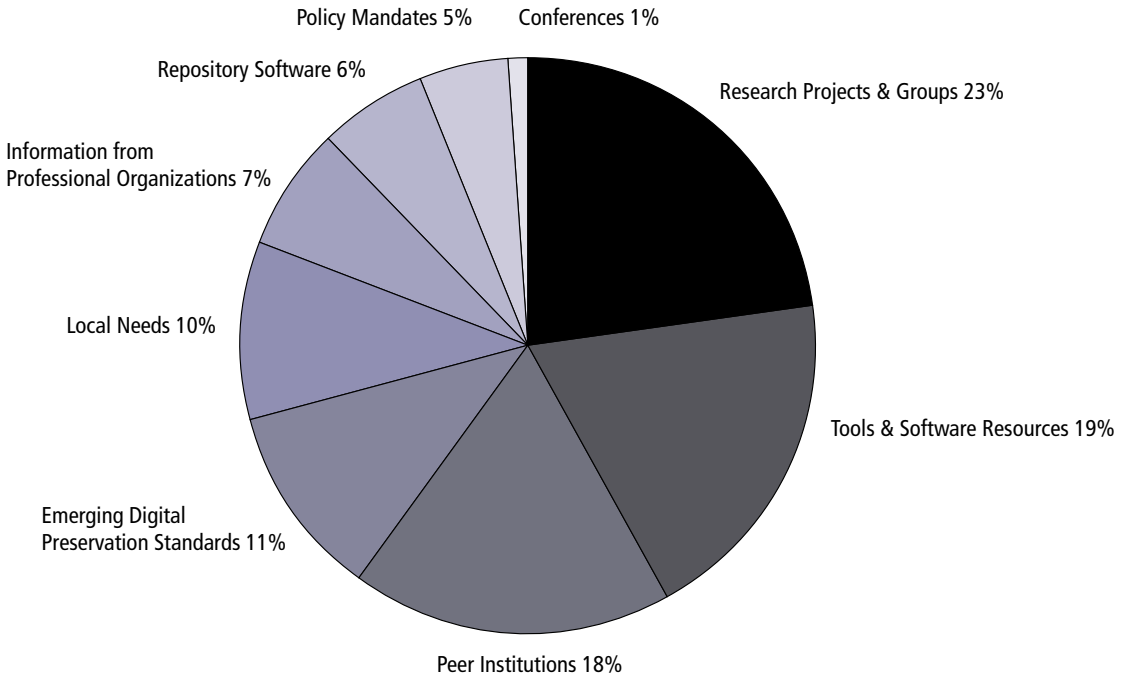
Almost all of the responding libraries (54 or 84%) are currently collecting electronic theses and dissertations. The majority also collect personal archives and institutional records and archives. Most of the others report they plan to collect these categories of materials. Twenty-one libraries collect research data and 28 others plan to collect it. Photographs, audio and video recordings, texts, and moving images are the most frequently collected media formats. About a third of the respondents collect websites, email, and databases; almost an equal number plan to collect these formats. While only six currently collect social media, 23 others plan to do so in the future.

Ingest Policies and Procedures

The majority of respondents (45 or 71%) have not developed gift/purchase agreement language that is specific to born-digital materials, but many are reviewing those agreements. Thirty-six respondents (56%) reported that they have developed ingest and processing workflows. An analysis of the comments indicates that a number of libraries are in the development phase. The comments also revealed a variety of models and/or examples the libraries have used in the development of workflows. These influences can be grouped into nine general categories as seen in the chart below.

Projects that influenced workflow development include the Personal Archives Accessible in Digital Media (PARADIGM) and futureArch projects at the University of Oxford's Bodleian Library, the AIMS project (Born Digital Collections: An Inter-Institutional Model for Stewardship) conducted by Stanford University, Yale University, University of Virginia, and University of Hull (UK), InterPARES, the British Library's Digital Lives project, the Tufts Accessioning Program for Electronic Records (TAPER) project, the European Union's Preservation

Influences on the Development of Ingest and Processing Workflows



and Long-term Access through Networked Services (PLANETS) project, and the Sustainable Archives & Leveraging Technologies (SALT) research group at the University of North Carolina.

Influential tools and software resources include Archivematica, the Duke Data Accessioner, digital forensics tools (including AccessData FTK Imager), file identification and validation tools (such as DROID and JHOVE), and the University of North Carolina's Curator's Workbench.

Respondents highlighted documentation made available by the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, the Digital Preservation Management workshop developed at Cornell University, the University of Illinois's IDEALS (Illinois Digital Environment for Access to Learning and Scholarship) repository, the California Digital Library's Merritt repository, Stanford's digital forensics lab, Emory University's Salman Rushdie collection, and Chris Prom's Practical E-Records blog.

Standards that influenced workflow development include the Open Archival Information System (OAIS) Reference Model, the PREMIS (PREservation Metadata: Implementation Strategies) metadata schema, the SWORD (Simple Web-service Offering Repository Deposit) protocol, and the BagIt specification.

Information provided by the MetaArchive, the National Digital Information Infrastructure and Preservation Program (NDIIPP), and professional journals, as well as the Digital Curation Centre's lifecycle model, influenced several respondents.

Perhaps as a sign of how workflows are tailored to fit local resources, some respondents cited DSpace repository software and CONTENTdm as influences on workflows. A few cited policy guidelines and mandates from parent organizations. Others mentioned Society of American Archivists and Midwest Archives Conference panel presentations on practical approaches to born-digital records, although no one mentioned conferences such as iPRES or the Personal Digital Archiving conference for which born-digital content is the specific focus.

While it appears that many respondents do not yet have well-established workflows for the ingest and processing of digital content, the majority are actively addressing the challenges of preparing born-digital content for long-term preservation and access.

Ingest Strategies

Seventy-seven percent of the responding libraries are ingesting born-digital records that are stored on legacy media. Almost all of them are storing the media "as is," and about half are collecting hardware that can retrieve data from those media. Fifteen libraries (25%) are outsourcing data retrieval and another 20 (33%) are planning to use that strategy. Only eight libraries are building new systems that replicate the function of the legacy systems. Other strategies include migrating content from legacy media to a storage location (described variously as "server storage" or "dark archives" space) and converting legacy born-digital content to "modern," "less proprietary," or "the latest usable" formats that include CSV files and PDF/A files.

Storage Solutions

The survey asked which kinds of storage media are used for ingest, processing, access, back up, and long-term dark storage functions. Most respondents use a combination of external media, network file systems, and local storage for all functions. Only 12 respondents (19%) report using cloud storage.

Local/attached storage (46 responses or 75%) and external media library (41 or 67%) were the most prevalent ingest solutions, followed closely by a network file system (35 or 57%). Other solutions include the DSpace-based commercial hosted Open Repository, the OnBase commercial enterprise content management system, and an institution's collection development instance of DSpace. One respondent stated that they are currently using cloud storage on a limited basis for ingest, and "plan to investigate its use for the other categories." Another belongs to a consortium that provides web-based ingest, processing, and access for ETDs, presumably including storage.

The most prevalent processing storage solutions are a network file system and local/attached storage, both at 43 responses (75%). External media library was a distant third. Other solutions were the same as for ingest: the consortium, the collection development instance of DSpace, and OnBase.

The most used access storage solution is a network file system (43 responses or 72%). External media library and local/attached storage each received 27 responses (45%). One respondent noted that they use Amazon Cloud and hosted Open Repository. Another uses a local DSpace instance, the California Digital Library's web archiving service, and a university system-wide open access repository. Other solutions include the use of a local implementation of a Fedora repository, YouSendIt online file sharing software in combination with e-mail, and shared IT servers.

The most common back up storage solution is a network file system (44 responses or 76%), followed by external media library (31 or 53%), local/attached storage (23 or 40%), and distributed systems (16 or 28%). Other solutions include a combination of Amazon Cloud and hosted Open Repository, the California Digital Library's Merritt Repository, redundant storage managed by campus and library IT, and physical tape storage.

Network file systems are used most for dark storage (26 responses or 52%), with distributed computing/storage systems second (19 or 38%). External media library and local/attached storage were not far behind at 16 and 14 responses, respectively. Other dark storage solutions include the California Digital Library's Merritt Repository, the Chronopolis digital preservation network, the Isilon commercial storage platform, redundant storage managed by campus and library IT, and virtual and physical tape storage. One respondent stated that rather than dark storage, their institution uses Fedora as an asset management system and copies files to "replicated storage for long-term preservation, with appropriate preservation metadata and restricted access."

Estimating Storage Needs and Costs

Twenty-six of the responding libraries (59%) estimate future digital storage needs and costs based on past and current usage and/or planned growth. Three noted that storage is allocated on a case-by-case basis. Some respondents have yet to implement methods of estimating storage needs and costs. Others are in the process of developing such methods.

Respondents described a variety of approaches to estimating storage needs and costs. One is conducting a longitudinal analysis of trends in digital storage growth. Another will scale future digital storage needs to the "development of campus department operations." Another currently uses costs of disks, storage devices, and backups as the basis for total cost estimates and is looking at moving to endowment-based storage cost models in the future. One respondent anticipates using the L.I.F.E. (Life Cycle Information for E-Literature) model developed by University College London (UCL) and the British Library for estimating curation costs, including the cost of storage.

One institution estimates space needs based on "past collecting volume + a 20% inflator + any known collections we anticipate receiving." Another estimates required storage needs based on average file size for a particular type of record and then estimates costs based on the current market value of storage, "usually at the TB level."

The most detailed response described the institution's attempt to estimate storage needs by tracking historical usage and growth, contrasting those with earlier projections, and categorizing data by type to identify growth areas. Thus far, the respondent observes that "consumption generally increases by a factor of 2 to 4 within a 12–18 month period," but any projection can change when unexpected projects or changes in the organization occur.

Access and Discovery

The survey asked which delivery methods the library uses to provide access to born-digital materials. Two-thirds of respondents provide online access to a digital repository system. Just under half provide in-library access on a dedicated workstation. Users who bring their PCs to 22 of the responding libraries can access born-digital materials stored on portable media. Eighteen respondents (28%) use third-party systems such as CONTENTdm, Archive-It, Dropbox, and YouTube to share materials with researchers.

There is not one, single repository system being used either to manage or provide access to born-digital materials. Most respondents use open source repository software for both management and access functions. Twenty-eight institutions report using secure file system storage to manage collections but only

ten use it to provide access. The results seem to suggest that access to collections is not as fully developed as the management of born-digital content.

The survey asked whether the institution is using different types of repositories for different types of born-digital materials. While 63% reported that they are, their comments indicate that they use different repositories for a variety of reasons, including media type (e.g., images, audio/visual materials, websites), record type (e.g., thesis and dissertations, faculty preprints), access and preservation requirements, and whether the material is digitized or born digital.

Ingest Challenges

The challenges related to the ingest of born-digital materials can be grouped into three broad categories: the difficulties associated with accessing information stored on legacy media and/or in obsolete file formats; the lack of policies, end-to-end workflows, and robust, integrated systems for digital object ingest; and the need to scale up to meet the increasing volume of born-digital objects needing preservation.

The challenges related to working with legacy formats and hardware were the most frequently cited ingest issues (43% of respondents listed file format or software obsolescence; 38% included legacy media or hardware). Donors, campus offices, and other records creators place their materials in a library or archives when they are no longer actively using them. As a result, libraries often receive storage media (punch cards, floppy disks, hard drives, CDs, zip disks, etc.) that are no longer accessible through current technologies.

Being able to transfer the files to appropriate storage is only the first step. The archivist then needs to be able to open them to assess their content. Obsolete file formats sometimes cannot be opened or executed using current software. Older versions capable of opening the files might require specific environments (operating systems and hardware) to run. Copyright restrictions and the terms of software licenses may make it difficult or impossible for staff to locate versions they can legally use. In addition, digital objects accessed through more modern systems often render differently than they did in their original environment. The formatting or appearance may be altered, and sometimes the behavior or even the actual content will change. Without the ability to access the content of older digital objects, it is difficult to determine which digital materials are most important and how best to allocate resources among collections. Given these challenges, nearly three quarters of respondents reported that their institutions store at least some of their legacy media as is, without transferring to new media or to server storage.

Collection donors have used a very wide variety of hardware and software configurations over time. As one respondent noted, "Each new collection seems to bring new technical issues that must be dealt with." In most libraries, it is unclear who should be responsible for developing technical solutions for accessing legacy media and obsolete file formats. This work is often outside the mandate of the information technology division and usually beyond the expertise of special collections staff. Some libraries and archives are creating "ingest labs" in house (the Bodleian Library, the British Library, Stanford, and the University of Virginia have working labs that serve as potential models). Others are outsourcing file recovery. An alternative file management strategy is to use a tool such as the Catweasel universal floppy disk controller, which is designed to connect legacy floppy disk drives to modern computer systems so that data can be read and written to floppy disks.

Interestingly, few respondents discussed challenges associated with complex digital objects (comprising more than one file and/or more than one file type), social media, digital objects stored in the cloud, websites, and networks of information, presumably, because most special collections and archives are just beginning to work with these types of digital objects.

The second category of ingest challenges relates to the workflows and systems needed to manage the digital objects once they are transferred off of their original carrier media. Maintaining privacy and providing adequate security topped the list of concerns. Respondents called for privacy and security policies specific to digital objects that address donor concerns and that insure compliance with university policies and federal and state laws. They noted the need for secure storage and networking and for tightly controlled access to files that contain personally identifiable

information. (See Kirschenbaum, *Digital Forensics*, pages 49–58 for additional discussion of privacy and security issues related to born-digital objects.)

Several respondents noted that archivists need to be able to dedicate more time to developing policies and conducting test pilots. The lack of clear policies and workflows can lead to inconsistent practices across collections and across the institution, and to inefficient resource allocation. Without consistent policies and procedures libraries cannot insure continued access to the born-digital objects. The PARADIGM project (Bodleian Library) and AIMS project both provide guidance in establishing policies and workflows. The BitCurator Project, led by the School of Information Science at the University of North Carolina at Chapel Hill and by the Maryland Institute for Technology in the Humanities at the University of Maryland, is building on these efforts. It will define and test a digital curation workflow, beginning at the point of encountering holdings that reside on removable media and ending with interaction with an end user.

The tools and systems used in the ingest process tend to be modular, and many were originally developed for use by other communities. For example, commercial forensics packages (which are very useful for browsing content and identifying personally identifiable information) were developed specifically for law enforcement. While the functionalities of these products have guided institutions in the development of workflows, they cannot be easily combined to meet the needs of the library and archives community. As one respondent noted, "There are several open-source and commercial products that can do pieces of the workflow, but as they are not designed to work together there are inefficiencies in stringing these workflows together." Another respondent added that "most ingest software is in alpha or beta release, with long-term roadmaps for future development." Early adopters and those libraries able to develop their own systems need to be comfortable with uncertainty and a certain amount of churn. Other archives are waiting for system development to catch up with their needs. Systems currently used include Archivematica, Rosetta, and the Curator's Workbench; others like Hypatia and BitCurator show potential for the future.

The final category of challenges related to ingest relates to the capacity needed to scale up workflows and systems to manage the flood of born-digital objects needing preservation. Respondents highlighted the need for sufficient storage space, adequate network capacity, increased staffing, staff training, automation of standard tasks, and enterprise-level systems. One respondent noted, "Our current archival storage was scaled to accommodate our analog to digital digitization program." It is more challenging to estimate the needs for born-digital special collections and archival materials: the timing for acquisitions can be hard to predict; the volume is not always known at the time of receipt (often because the digital objects are on legacy media); the formats often vary widely; and it is often unclear which materials will need to be restricted (because the files cannot be accessed before receipt due to media or format).

Storage Challenges

The challenges related to storage systems can be separated into three major areas: systems limitations, organizational challenges, and insufficient resources (i.e., not enough available space and high storage costs). The challenges surrounding systems limitations were divided between the need for preservation-quality infrastructure and the need for security for and access to the materials themselves. Organizational challenges fell into three categories: policy and planning, gaining and retaining sufficient staff and skills, and managing the organizational structure (from the department up to the entire organization) while maintaining effective coordination between all the stakeholders. One set of concerns about sufficient resources represents two sides of the same coin: insuring adequate file storage space and its cost. Other challenges related to storage space include the difficulty in estimating and predicting capacity needs. One comment that summarizes the issues well indicates that storage needs for born-digital records should not be only the responsibility of the library and archives: "Future storage needs for large-scale ingest of born-digital special collections materials will probably be integrated into university-wide planning for digital repositories, a digital asset management system, and networked storage and continuity planning."

Access Challenges

The biggest access and discovery challenge, described by 32 respondents, is the sensitivity of materials—concerns about copyright, confidentiality, privacy, intellectual property, and personally identifiable information. The second biggest challenge is IT infrastructure, or rather, the lack of it (28 respondents). Particular concerns in this area include user interface, the need to integrate multiple systems, and the ability to handle very large files. Other significant challenges are the need to develop policies, processes, and tools for arranging and describing born-digital materials in ways that make them most accessible, including the integration of description for digital and non-digital materials; rights management (restrictions specific to users rather than materials); and staff time and skills. Interestingly, *time* was twice as much of a concern for respondents as *staff skills*. This makes sense as more professionals are assigned responsibility for these materials and go on to develop the necessary skills, but *staff* may still mean the only person, or one of a very few, responsible for managing these types of materials at their institutions. The remaining concerns included metadata standardization, differing levels of donor restrictions and how to apply them in an online environment, format standardization and migration, and institutional support (including funding).

Respondents' concerns grow even more complex when restrictions on sensitive materials (those subject to copyright, confidentiality, privacy, and intellectual property concerns) are combined with rights management by user group and donor-imposed limitations on access, because each of these types of restrictions can vary from case to case. Reference desk staff have dealt with the complexity of access restrictions in face-to-face transactions for decades, but libraries lack automated systems that can do the same during online transactions where staff are not there to intervene.

Respondents' comments on registration procedures highlight the nature of this challenge. Most institutions that provide access to born-digital materials are either doing so in their reading rooms and following standard reading room registration procedures or are providing access to the materials online with no registration procedure. These limited approaches are directly linked to the second biggest access challenge for respondents, the lack of a fully developed IT infrastructure for delivering born-digital materials to researchers. Other technology concerns include user interface design, the need to navigate multiple disconnected systems, and problems supporting large file sizes.

Providing access to archival materials is, of course, dependent on appropriate arrangement and description, and so it should be no surprise that many respondents stated a need to further develop policies, processes, and tools for arranging and describing born-digital materials in ways that make them most accessible, including the integration of description for born-digital, digitized, and non-digital materials.

The survey results indicate that our profession is moving towards a higher comfort level with the standardization of both metadata and file formats. Furthermore, institutional support is a challenge at only three institutions, which would seem to illustrate administrators' growing understanding of the need to support access to born-digital materials. Possible areas for future research include the use of analytics and user studies to track the quantitative and qualitative aspects of access to these materials by off-site researchers and the challenges of providing not just basic access but value-added reference services to those users.

Privacy Concerns

The survey team was surprised that most respondents did not address the potential institutional liability posed by personally identifiable information (PII) within born-digital materials, beyond the imposition of access restrictions. (PII includes information such as social security numbers, credit card numbers, logins, passwords, PINs, and medical and financial records.) Seventy-one percent of respondents indicated that their gift agreements did not include language that acknowledged born-digital materials. While ownership transfer, copyright, and some standard restrictions can be handled through the traditional deed of gift, gaining permission from the donor to use forensic tools that allow recovery and review of deleted files while searching for PII is not a standard option. Since such actions might alter donated files or uncover files not intended for transfer, requesting permission through

the gift agreement or some other policy document is highly recommended.

While 71% of respondents have policies regarding whether files with PII should be retained with restrictions or destroyed, only 43% have policies indicating whether born-digital materials can be made available for research use before they are screened for PII. One respondent's comment that "all special collections materials have personally identifiable information (PII)" is quite true. However, paper-based collections have always benefited from security through obscurity. There is no fast or easy way to uncover social security and credit card numbers in paper-based collections. With born-digital records, on the other hand, there are many tools available that can search and locate PII, even in deleted or hidden files. Such content, improperly managed, not only puts the file creator at risk, but also may be in violation of an institution's security and privacy policy for this type of information. Eighty percent of respondents indicated that they do not have a written PII policy. Greater security is needed for unscreened born-digital records, especially if they are stored on networked servers.

Conclusion

The responses to this survey indicate that many ARL libraries and archives have begun working with born-digital materials in their collections, despite the fact that enterprise level systems and best practices for managing these materials in an archival setting are still in development, and despite concerns that they do not have the resources to scale their work to meet current and future demand. This willingness to experiment, to learn new skills, and to seek to understand the scope of the issues is building expertise within the library and archives profession, and has insured access to some born-digital holdings, at least in the near term. It also signals a shift from a wait-and-see attitude to a more empowered something-is-better-than-nothing approach to managing born-digital materials.

Respondents identified the following as critical for transitioning their work with born-digital materials from projects to programs:

- Collaborative solutions for dealing with hardware and software obsolescence.
- More, and more appropriate, storage for born-digital materials (long-term, authenticated, secure, verified, backed-up, and geographically distributed). As one respondent noted, "Archives are guaranteed preservation only if stored on enterprise data storage."
- Automation of as much of the workflow as possible.
- Asset-level access control to enable tiered access to restricted records.

Many institutions are working with digitized content or licensed digital content and are only now beginning to explore the ways in which born-digital, primary-source materials may be different. For example, it is difficult to estimate storage needs for born-digital primary sources stored on legacy media prior to accessioning and processing them. Privacy concerns are magnified when large bodies of easily searchable digital material may contain personally identifiable information. The workflows and infrastructure built for digitized content are often insufficient for born-digital primary sources.

While some special collections rely on a single staff member to manage all aspects of preserving and providing access to born-digital materials, more frequently staff from special collections, library IT, digital repositories, digital curation, and other areas work together to ingest, appraise, describe, preserve, and provide access to this content. The distributed nature of this model allows the library to leverage existing expertise, but it may also mean that no one has the big picture. These situtions make it difficult to track the resources needed to manage the materials—which then makes it difficult to estimate current and future costs. Distributed responsibility can also threaten the long-term survival of the materials, either when no one feels empowered to make decisions or when someone makes decisions without having all of the relevant information. Staff need models of existing

teams that describe how responsibilities are assigned and decisions are made collaboratively.

Survey responses indicate that best practices will take some time to develop: infrastructure, systems, and tools are in development; libraries continue to experiment with organizational models to find those that will be most effective to manage born-digital, primary-source materials; and the variety of record formats continues to grow. While several libraries and archives have developed scalable solutions that work within their own context, few of the solutions developed to date have been transferable between institutions. Waiting for time-tested systems and practices, however, is not an option. For now we need to settle for "good enough" practice and continue to invest time and resources in developing systems and workflows that will prevent a "digital dark age" for the first part of the 21st century.

SURVEY QUESTIONS AND RESPONSES

The SPEC survey on Managing Born-digital Special Collections and Archival Materials was designed by **Naomi Nelson**, Director of the David M. Rubenstein Rare Book & Manuscript Library, and **Seth Shaw**, Electronic Records Archivist, at Duke University; **Cynthia Ghering**, director of the University Archives and Historical Collections, and **Lisa Schmidt**, Electronic Records Archivist, at Michigan State University; **Michelle Belden**, Access Archivist and IT Coordinator for the Special Collections Library, **Jackie R. Esposito**, University Archivist and Head, Records Management Services, and **Tim Pyatt**, Dorothy Foehr Huck Chair and Head of the Eberly Family Special Collections Library, at Pennsylvania State University; and **Nancy Deromedi**, head of the Digital Curation division, and **Michael Shallcross**, Assistant Archivist, in the Bentley Historical Library at the University of Michigan. These results are based on data submitted by 64 of the 126 ARL member libraries (51%) by the deadline of March 26, 2012. The survey's introductory text and questions are reproduced below, followed by the response data and selected comments from the respondents.

The 2010 OCLC Research report, *Taking Our Pulse*, listed management of born-digital materials as the biggest challenge facing libraries, special collections, and archives, after space and facilities concerns. Over the last decade the materials acquired for our libraries, archives, and manuscript collections were very likely created as digital objects that may or may not have analog surrogates. If modern special collections and archives are to stay relevant and continue to provide access to unique and authentic records, ARL libraries need to manage and preserve born-digital materials, which for the purposes of this survey include institutional records, author's drafts on floppy discs, digital photographs and moving images, and electronic theses and dissertations, among others. It excludes commercial products such as e-journals.

This survey explores the tools, workflow, and policies special collections and archives staff use to process, manage, and provide access to born-digital materials they collect. It also looks at which staff process and manage born-digital materials and how they acquire the skills they need for these activities, and how libraries have responded to the challenges that managing born-digital materials present.

BACKGROUND

1. Does your library currently collect and manage born-digital materials? N=64

Yes	59	92%
No, but we plan to	5	8%
No, and we have no plans to	0	—

STAFFING

If collecting born-digital materials is in the planning stages, please answer the following questions to the best of your ability based on plans at this time.

2. Please indicate how many staff are (will be) charged with collecting and managing born-digital materials at your library. Include both the number of FTEs and number of individuals. N=60

	Minimum	Maximum	Mean	Median	Std Dev
FTE	0.10	60	4.73	2.00	9.25
Individuals	1.00	48	6.64	5.50	7.75

Please enter any additional comments you have about the staff who collect and manage born-digital materials. N=47

Special Collections/Archives

All curators and archivists could potentially collect born-digital materials, so I'm including all of them!

All staff who are currently responsible for paper records will/have responsibilities for managing born-digital materials.

Digital Records Archivist (full time), University Archivist, and two curators.

Four full-time professional curators collect born-digital materials along with analog materials and a wide variety of other duties. We have no dedicated field collectors.

In addition, we have one student (.25 FTE) and another part-time intern.

Our University Archivist is our point person for born-digital material.

Right now we have a full time Digital Archivist in our Film and Media Archive, and a staff member in University Archives who has 50% of her job designated for digital collections. I would anticipate needing to add more staff capacity for this in other units of the department.

The breakdown of FTE hours to staff above reflects one person, the Digital Archivist, working solely on collecting and managing born-digital materials as well as 5 other staff members spending part of their time to reach the equivalent of 2.0 FTE. These numbers do not include those outside of the special collections unit, such as in preservation, technical services, and IT units, who help to develop systems like repositories and catalogs that help manage these materials.

There are currently two individuals in our Special Collections Department who play (and will continue to play) a key role in collecting and managing born-digital materials, but given the primary responsibilities of these individuals, their time (collectively) does not constitute even a single FTE.

Digital Curation/Repository Staff

Although we are collecting and managing born-digital materials, there are not specific job descriptions within the archives that are related to such activities. Working with born-digital content is under the purview of archivists and select staff. The library does have a digital preservation coordinator but that position addresses only select parts of managing born-digital content.

Archives and Digital Collections both expect to have a role in managing these materials.

We have one professional managing our institutional repository.

We have three collection areas for our repository where we collect and manage born-digital materials. The core collections manager works with scholarly resources (ETDs, faculty deposits, and general collections), as well as providing oversight for all born-digital collections. The Digital Archivist oversees Special Collections/University Archives collections. The research collection manager manages research data. The research associate (new position) assists him with faculty outreach and collection building. Research data is a rapidly growing area for born-digital materials. The Digital Data Curator sets digital preservation standards and manages the ingest, durability, and security of all digital collections and the Digital Projects Coordinator oversees the workflow of all digital collection building. None of these positions works exclusively with born-digital materials, but all work with some percentage of born-digital materials. We estimate that born-digital resources represent approximately 20–25% of our current collection ingest. We have many positions that create metadata and develop tools for the repository but they are not specifically tasked with collecting and managing born-digital materials.

Various units/unit not specified

Arts Library has 2 FTE plus 2 students at 25%; Research Data Curation figures include both librarians who work on collections and technology staff who build the storage and discovery applications.

Currently .5; plan to hire one FTE this year, and train an existing FTE the following year. So the above reflects this.

Includes Electronic Records Archivists, Digital Curation Librarian, IT Staff.

Institutional Repository (IR) Coordinator, Digital Humanities Librarian, staff in the Digital Development and Web Services Unit, staff in the Digital Library Center within the Digital Services and Shared Collections Department, and faculty and staff in the Special and Area Studies Collections Department.

It is difficult for us to break out FTEs for this work, as it is generally integrated with other work. For example, our Records Manager is responsible for working with digital content from the university; our Digital Initiatives Librarian works with born-digital as well as digitized content, etc. We also have three people outside special collections areas that spend a portion of their time working with electronic theses and dissertations; they are not included in this figure.

Most of the staff involved with these projects participate less than full-time. Group includes librarians, technicians, developers, and project managers.

Nearly all staff members have some responsibility for digital materials, but only as part of their job. Estimate is therefore very rough. One staff member is oriented predominantly toward digital.

No one staff member is charged solely with this responsibility. Rather, all professional staff who have a role in acquiring new collections also have the responsibility to undertake these tasks.

One FTE is for the Director of Research Systems Development, who is not a Special Collections staff member but manages the Institutional Repository where the dark archives are located. Another is for non-Special Collections staff who manage the instance of CONTENTdm, which includes accessible born-digital materials.

One staff member is tasked with developing and maintaining the born-digital accession workflow process, all staff work with born-digital content in some capacity in the arrangement and description process.

Only one of these positions is devoted full-time to managing/collecting digital special collections materials.

Over the next few years, we hope to increase to 3 FTE (2 FTE permanent staff and 1 FTE project staff or interns).

Responsibility for collecting and managing born-digital materials is currently shared by librarians and archivists with responsibilities for special collections, university archives, geospatial data, ETDs, cataloguing & metadata. The library will soon hire a Digital Special Collections Librarian who will take the lead on collecting and managing born-digital special collections. This will lead to a higher FTE number than reported here.

Staff include University Archives personnel, the Faculty of Medicine Archivist, Libraries Collections Management personnel, the University Records Manager, Libraries IT personnel, and contract metadata technicians.

The Libraries have recently reorganized, still in process of figuring this out.

The numbers above speak to departments with particular responsibility for the management and long-term support for digital files, not for the collection development aspect. Collection development of born-digital materials will be carried out by curatorial and archivist staff not reflected in these numbers.

The staff are not dedicated only to this activity but it falls under the scope of other archival work.

There is no one person who does this full time. Everyone involved is focused on this issue as part of all of their other duties.

These individuals are not devoted exclusively to born-digital materials, but born-digital materials will invariably be part of the collections these individuals acquire, organize, preserve, and describe. The Head of Special Collections and Archives collects archival materials, which increasingly include born-digital materials. The Archivist is responsible for arranging and describing archival materials, including born-digital materials. The Digital Project Specialist assists the Head of SCA and the Archivist in acquiring, storing, preserving, describing, migrating, and providing access to these files. The Digital Projects Specialist administers the various digital repositories that preserve and provide access to these materials as well.

These numbers are very difficult to accurately compile. Most staff members do not have hard time allocations for born-digital materials. Most staff members do not have explicit job descriptions regarding born-digital materials. Also, those staff who do have allocations or explicit job descriptions may also be responsible for other tasks.

This includes metadata experts, digital curation staff, and repository services.

This question is difficult to answer with any accuracy. We currently collect very little born-digital materials and we have no one individual that is dedicated to this task or will be dedicated to it within the foreseeable future. Almost all of our special collections receive a small amount of born-digital materials and therefore the staff that is "charged" with managing and collecting the materials are no different than those that collect our paper-based materials.

This will be part of everyone's responsibilities.

This will become a component of the work of each of our four archivists responsible for acquiring all archival materials, regardless of format.

Two staff in Special Collections collect and manage born-digital content. In our Scholarly Publishing and Data Management team we have five individuals who are managing digital content.

We are at the earliest stages of formulating a formal strategy for ingesting born-digital Special Collections content. The figures provided above are a best guess as to how many people may be involved and makes the following assumption: University Archives staff member (0.25 FTE), Manuscripts Division Staff member (0.25 FTE), Programmer/Analyst (0.25 FTE), Metadata Specialist (0.75 FTE).

We do not have a full-time person, and since we are doing this on an ad-hoc basis now, the quarter FTE is really just a guestimate. I, through my curatorial duties, and our digital services librarian handle it.

We have 60% time of two people (both professional staff), roughly half the time of one software engineer, and a small amount of time from a core services/Mac specialist team member.

We have added these duties to the work of the Technical Services Archivist.

We have numerous staff who are involved with managing or collecting born-digital, but none of these positions are dedicated full time to the activities.

We have three FTE staff who have some responsibility to manage born-digital content in some capacity in their job descriptions: Archivist, Digital Projects and Outreach Digital Assets Librarian (currently conducting job search to fill this role), Digital Initiatives Librarian.

Work falls into three general areas: web archiving, electronic archives, and audio and video oral history interviews.

3. Please indicate which strategy your library has used or plans to use to address staffing needs for processing and/or managing born-digital materials. Check all that apply. N=63

Add responsibilities for born-digital materials to current positions	59	94%
Create new staff positions	29	46%
Recast an existing position dedicated to managing born-digital materials	23	37%
Hire interns for born-digital materials projects	15	24%
Hire consultants/contractors for born-digital materials projects	9	14%
Other staffing strategy	13	21%

Please describe the other staffing strategy. N=13

Also hoping to work with staff from larger library who are charged with managing born-digital NON special collections.

As the library makes strategic hires in other areas (e.g., Digital Repository Coordinator), we will attempt to include electronic records expertise in the desired skill sets.

Budget for staff in grant-supported digital preservation projects.

Capitalize on existing expertise to leverage capacity where it exists.

IT Specialists are being used. Currently Still Pictures just add responsibilities for born-digital materials to current positions. Electronic records preservation staff directly contribute to the Electronic Records Archive activities. Considerable workload has increased pertaining to system design, development, testing, and analysis.

May link up with other units in the Libraries or beyond.

One librarian has had his position expanded to include acquisition and management of gaming collections.

Recast an existing position to include managing born-digital materials.

The e-records/digital resources archivist position that is split with the office of the CIO does investigate tools and provides advice to curators and curatorial staff in more effective ways of accessioning and managing born-digital materials.

Trained high-level students dedicated to born-digital projects (4 students).

We expect there will be a need to create new positions and/or recast existing positions for managing born-digital materials.

We have a newly created digital resources library unit that will work with born-digital.

We have created new staff positions. The Digital Archivist, a position in the Special Collections/University Archives Unit, manages digital collection building for that unit. The Digital Project Coordinator position was created to add oversight and accountability to the many digital collection-building projects underway. We repurposed a position to become the Digital Data Curator, and another position to become the repository Digital Collections Manager. We added responsibilities for digital collection management to an existing professional position, the Social Sciences Data Librarian. We are actively working with other library liaisons to build digital collection building into their service repertoire. We hope to add an electronic records management position in the next 1 to 2 years in Special Collections/University Archives, to manage the official electronic records of the university. This position will work with an ER system, rather than with the repository although we anticipate sharing born-digital materials from the ER system that have scholarly value and are available for open access use across the two systems.

4. What opportunities does your library pursue (or plan to pursue) to increase staff expertise in managing born-digital materials? Check all that apply. N=64

Conference attendance	58	91%
On-the-job training	58	91%
Training provided by professional organizations	57	89%
SAA workshop/Summer Camp attendance	41	64%
Independent study	39	61%
Local courses in computer or digital technology	21	33%
Training provided by vendors	15	23%
Rare Book School attendance	14	22%
ARMA workshop attendance	8	13%
Other opportunity	9	14%

Please specify the other opportunity. N=9

Anything we can get!

CURATEcamp (curatecamp.org)

Mentoring

METRO, NYART; also engagement with groups such as PASIB, LibDevConX; webinars from DuraSpace and NISO; discussions with colleagues from other institutions, especially Cornell.

Peer-to-peer on-the-job training

Regional meetings such as the Northwest Archivists, also the Northwest Digital Archives part of the Orbis Cascade Alliance.

Vendors in this case refers to training conducted by our prime ERA contractor (Lockheed Martin, in 2009 and 2010). On the job training: "We all help each other as we learn about the utility of new software tools, skills, and techniques in processing digital images." Several staff have participated in electronic records management courses conducted in field locations.

We are currently teaching an in-house course on research data management for library faculty liaisons, metadata librarians, and library technologists. The emphases are on understanding the nature of research data, metadata, rights and privacy, and data curation, with a goal to building project teams to work effectively together on research data management. This is not exclusive to born-digital content, but we are finding most of our research data is born digital.

Workshops/Institute such as DigCCurr; CurateCamp; Digital Preservation Management Workshop.

BORN-DIGITAL MATERIALS COLLECTED

5. Which of the following categories of born-digital materials does your library currently collect or plan to collect? Check all that apply. N=64

	Currently Collect	Plan to Collect	N
Personal archives such as email, photographs, documents, calendars, etc.	46	13	59
Organizational or institutional archives	41	18	59
University/institutional records	44	13	57
Electronic theses and dissertations	54	4	58
Research data	21	28	49
Non-commercial e-only publications	30	11	41
Learning content and course management systems	8	11	19
Other category	18	—	18
Number of Responses	64	41	64

If you selected "Other category" above, please briefly describe the other category of born-digital materials your library currently collects or plans to collect. N=17

Publications/related materials

Academic department newsletters, documentation from university research centers

Commercial e-only publications through copyright deposit and purchase/subscriptions. As part of our manuscript collection efforts we do collect materials from individuals. Donations, gifts, and exchange items are another channel for acquisition.

Currently collect commercial e-only publications.

Digitized books, campus websites, campus journals, etc.

AV/Research data

Faculty portfolios, campus video productions, oral histories

I included theses, research data, and content management systems even though they are the province of the larger university library system, not special collections. Our collections in all areas are very small, perhaps just a few items as test cases, and in some cases, simply being stored on drives counts as collection.

Oral histories

Oral history video/audio interviews

Research data is still an unclear area, in part simply defining "research data" is difficult. Re: course management material, that area is also unclear.

Scholar curated archives and research materials in the humanities, often referred to as capta instead of data to recognize that the data is not discrete/quantitative, but data as it is captured/presented; born-digital materials from other GLAM (galleries, libraries, archives, and museums); oral histories; web archiving.

Born-digital audio and video oral histories created at the university. May collect social media from institutions/ individuals whose archives we hold.

ETDs/Student works

Other types of student works

Outside of ETDs what we collect born-digital is currently minimal—some university records, an e-pub, and undergraduate honor theses. However, ETDs and honor theses do not come under the curatorial purview of Special Collections and Archives.

Other

Maps, catalog indices

Thematically related websites, e.g., in the areas of human rights, historic preservation/urban planning, religion/theology, and personal websites for individuals whose archives we hold.

We are the custodians of the electronic Presidential records transferred to NARA at the end of each administration. While Presidential records will remain the vast majority of our electronic holdings we also have small volumes of personal electronic records that will need a more systematic management approach in the coming years.

We currently access records and fonds that include born-digital material, mostly on media such as hard drives, floppy disks, and CDs.

Websites, blogs, etc.

Additional Comments

Content support born-digital publications and artworks.

Electronic theses and dissertations, research data, and non-commercial e-only publications are handled through the Institutional Repository, which is not part of Special Collections.

Research data: initiatives in this area are currently managed by the Research Data Management Service Group within the University Library, not within special collections units.

6. Which of the following types of born-digital materials does your library currently collect or plan to collect? Check all that apply. N=64

	Currently Collect	Plan to Collect	N
Audio recordings (including podcasts)	49	14	63
Photographs	53	9	62
Moving images	42	19	61
Video recordings	48	13	61
Texts (such as unstructured office documents)	43	12	55
Institutional websites	24	28	52
Email	23	23	46
Databases	21	20	41
Other types of websites	18	20	38
Geographical Information Systems (GIS) data	19	12	31
Social media (e.g., institutional Facebook pages, Twitter accounts)	6	23	29
Executable files	14	4	18
Enterprise systems data	2	13	15
Computer games	5	5	10
Other type	6	6	12
Number of Responses	62	49	64

If you selected "Other type" above, please briefly describe the other type(s) of born-digital materials your library currently collects or plans to collect. N=12

Currently Collect

Excel spreadsheets

Illustrator, InDesign files (design production files), iMovie project files

Music scores. Outside of Special Collections, the institutional repository collects born-digital moving images, video recordings, databases, texts, executable files, and GIS.

Posters and other graphic materials in digital form

Serials and monographs

Plan to Collect

Any materials we've collected in the past that are now in an electronic format.

Architectural design files (e.g., CAD)

Because we collect archival material from a variety of external people and organizations, it is difficult to predict exactly what formats of born-digital material we may acquire in the future.

Much of this territory is still in discussion.

Oral histories

Research datasets

Additional Comments

Research data in various formats (.csv, XML, tab delimited, etc.)

INGEST POLICIES AND PROCEDURES

7. Has your library developed language for your gift/purchase agreements that is specific to born-digital content? N=63

Yes	18	29%
No	45	71%

Answered Yes

Archives does not distinguish between traditional and digital records in this document. Our copyright statement is inclusive of born-digital and digitized material.

Gift agreement includes a note regarding electronic records and requires the donor to agree that there are no other copies of the e-records available elsewhere.

Gift agreements acknowledge the possibility despite the fact that it is not clear to me or our counsel why special mention is necessary.

Language for gifts/purchases that include digital content is developed on a case-by-case basis.

Only for ETDs. Nothing yet for general donations of born-digital materials.

Only when applicable to the collection.

The language is an optional add-on to the existing donor agreements.

We have specific language in some of our agreements, but we have not yet standardized the language or created a set of standard language that can be used for born-digital collections. We are working on standardizing our approach.

Answered No

But we are currently drafting such language.

In general, we feel that our regular agreements will cover most issues related to born-digital content. In the cases where this is not the case we will customize these agreements to address specific issues.

In process of developing.

No, we are "media independent" in our deeds.

Our current licenses cover both digitally converted and born-digital content.

Our gift agreements already encompass most issues pertaining to the management of born-digital content, but not language specific to born-digital content.

Our intake of born-digital materials is still small and is often linked with traditional analog collections.

Revision of gift/purchase agreements will be one focus area of the new Digital Special Collections Librarian.

So far, only ad hoc language for a single born-digital archival collection.

The agreement language has not yet been finalized, but it will be updated to reflect areas that were not as relevant to paper materials.

This does need to be addressed for reasons of preservation/migration and also online access.

We did develop a special deed of gift for project involving solicitation of images and other born-digital content from the community, because we were concerned that the casual way in which we were likely to receive content would reduce our chances of getting donors to sign our standard form. Whether born-digital content is more likely to be donated casually (i.e., the digital equivalent of being left on the doorstep) is hard to say at this point; if it is, then this was an adaptation to that quality.

We have amended some specific deeds of gift to allow for electronic publication of born-digital content, but this is very rare so far.

We have begun the discussions to consider how to include those revisions.

We will be addressing this in the near future.

Working on gift agreement to include all types of media, including born-digital content.

8. Has your library developed workflows for the ingest and processing of born-digital materials? N=64

Yes	36	56%
No	28	44%

If yes, please briefly describe any models or examples you found most helpful as you developed your workflow.

Models/Examples

Currently using Google Spreadsheet APIs for ingestion, but interested in approaches such as SWORD and BagIt.

Digital Curation Centre lifecycle model

Examples provided by the Digital Preservation Management Workshop at Cornell.

Inspiration for this process came from UT Austin's digital preservation work (https://pacer.ischool.utexas.edu/handle/2081/21808) and Stanford's digital forensics work (http://lib.stanford.edu/digital-forensics), as well as the steps and validation processes in Archivematica and Duke Data Accessioner (though we are not currently using these tools).

Model(s) that were helpful to Preservation staff in developing workflows for ingest and processing born-digital records was the Open Archival Information reference model and the Digital Curation Center model.

Models we have used include documentation from Stanford and the Bodleian, as well as microservices as described by Archivematica

OAIS, IDEALS

PARADIGM; Existing accessions process for analog materials.

The AIMS project, specifically Stanford's work on digital forensics.

The most useful examples are real-world use cases for the full process of how to support the ingest, normalization, preservation, and access of born-digital files within a full repository and digital library system.

We are in the process of creating our workflows for ingest and processing. We have some workflows now that will soon change as our repository infrastructure evolves. Models we have used include documentation from Stanford and the Bodleian, as well as microservices as described by Archivematica.

We have studied and learned from the Duke Data Accessioner, the PARADIGM project in the UK, the OAIS model, and professional literature in developing our workflow.

We participated in the AIMS project and have developed a model for ingest from media following a forensics model. We are also utilizing forensics tools to enable arrangement and description.

We referenced many models while developing our own. Primarily, our workflow is based on the work we undertook as part of the Born-Digital Archives: An Inter-Institutional Model for Stewardship (AIMS) grant. That project in turn synthesized many research projects and workflows, but was heavily influenced by the following projects and tools: PARADIGM, OAIS, futureArch, Digital Lives, InterPARES, TAPER, SALT, the work of Chris Prom on his Practical E-Records

blog and report to, Archivematica, Curator's Workbench, work on the Salman Rushdie Papers at Emory University, PLANETS.

Work by Seth Shaw and Ben Goldman in conference presentations on practical approaches to born-digital collections at the Midwest Archives Conference and the Society of American Archivists. Duke Data Accessioner. Chris Prom's blog. CIC electronic records policy guidelines. MetaArchive and ICPSR's guidelines on development of digital preservation policies. Specifications of processes/tools/procedures from Archivematica, California Digital Library's Merritt, MetaArchive, etc. Publications from NDIIPP, ICPSR, InterPARES, PREMIS, etc.

Workflow Descriptions

A procedure to receive digital images and assign file names according to local directory needs is in place. Scripts for ingesting ETDs from ProQuest. Ad hoc scripting to structure and ingest research data.

Currently the workflow is very straightforward and is intended to protect the records against loss due to failure of the information carrier.

One principal driver for us was compliance with the requirements of the Presidential Records Act (PRA). The PRA gives the Archivist legal custody of all Presidential records at the point of an administration transition. The PRA also obligates NARA to respond to access requests to those records immediately after we receive custody (public access requests begin five years after transition; in the first five years we respond to special access requests). To meet both these circumstances our workflows have to account for the ingest of a large volume of holdings in as short a time frame as possible while giving us search and access capabilities to support asset-level review and production of copies of the electronic records for external requesters. Model(s) that were helpful to Preservation staff in developing workflows for ingest and processing born-digital records was the Open Archival Information reference model and the Digital Curation Center model. In Still Pictures, we have a multi-page set of basic instructions that cover what processing is, but essentially we: 1) obtain the digital images from the agency usually by downloading onto media for transfer to NARA. 2) Once here we make a copy for OPA processing. 3) Processing accession for ERA involves reviewing images to delete those that are temporary; ensuring unique filenames for images; appending our RG and series designations to each digital image; when images do not have captions, appending whatever information is available to each image in a folder; reviewing the metadata to make sure there is a link to the individual images; if caption information is in header, copying that out into a separate text file if needed. Depending on the condition of the accession, there may be many other processing steps needed to make it ERA and OPA ready. 4) Go thru the laborious process of ingesting the accession into ERA. 5) Complete processing for OPA and work with NPOL to get the images uploaded to OPA for reference use.

Our process is still being developed and tested. Currently it includes the following elements: Capture, metadata/content extraction. FTKImager to capture disk image, generate disk/file level metadata and checksums, and extract content directory from disk. BASH shell script to combine and organize disk image and metadata files. File Characterization/ Normalization JHOVE and/or DROID for characterization/validation. FileMerlin to convert/normalize legacy text files. Adobe Acrobat to migrate text files to PDF/A. Appraisal, organization, and description (akin to traditional archival processing). Human uses Excel spreadsheet to record appraisal decisions, organize content, and enter descriptive metadata. Ingest XSLT used on Excel spreadsheet to package the digital files and create Dublin Core .xml metadata files for ingest into our DSpace repository. Command line batch ingest to DSpace.

Our workflows are not specific to born-digital materials. For electronic records management, we have record schedules and retention policies that apply equally to analog, digitized, and born-digital records. For the digital repository, we utilize a workflow management system that enables us to establish collections, develop and document master file formats, validate and document technical characteristics of files, develop metadata, attach digital files to metadata, and

create and ingest METS packages. None of this workflow is specific to born-digital content but works equally well with digital and born-digital content.

We have different workflows for different content types. All workflows are preliminary and evolving.

We have implemented Archivematica as a key element in our accessions workflow.

We have localized workflows for some of our materials (e.g., EDTs), and are looking at methods for managing ingest in a more distributed or decentralized manner.

Workflow for our theses & dissertations: students submit electronically through a web form in our DSpace repository and there are two levels of validation by people afterwards within the repository before it is published. We used workflow models from other universities when setting our own.

Workflows in Development

In process of developing.

In process of development as library; VRA workflow model used by art image library.

There are some workflows in place, however, they are in the process of being reviewed and modified/expanded.

This is a work in progress.

This is in the midst of change. Based on a preservation repository model. Current challenge is the model for collections being processed and what to do until they are processed.

Training the campus photographers to add some of their images into a CONTENTdm system hosted by the library.

We are currently in the process of creating these workflows and hope to have them implemented by the end of the 2012 calendar year.

We are currently in the process of creating these workflows with a vendor.

We are experimenting with ingest into DSpace. Very early stages.

We are in the midst of developing proper ingest procedures: some parts worked out, some not so much.

We are in the midst of developing such workflows. However, we are building the infrastructure to support these initiatives from the ground up and do not anticipate being able to ingest significant quantities of born-digital content for at least a year.

We are in the process of developing workflows.

Workflows and policies still in development. Waiting for sufficient, *secure* storage infrastructure.

Workflows are in development.

9. Does your library currently ingest born-digital records stored on legacy media? N=64

Yes	49	77%
No, but we plan to	12	19%
No, and we have no plans to	3	5%

10. Which of the following strategies does your library employ when ingesting born-digital records stored on legacy media? Check all that apply. N=60

	Current Strategy	Planned Strategy	N
Storing legacy media as is (without transfer to new media or server storage and/or keeping it with analog collection)	47	1	48
Developing a collection of legacy hardware that can be used to retrieve data from legacy media (e.g., 5.25" floppy drives, zip drives, etc.)	27	10	37
Outsourcing the process of retrieving the data from legacy media	15	20	35
Building new systems that replicate the function of the legacy systems (e.g., emulation, virtual systems)	8	9	17
Participating in a collaborative that is developing a collection of legacy hardware	3	8	11
Other strategy	13	9	22
Number of Responses	57	32	60

If you selected "Other strategy" above, please briefly describe the other strategy(ies) your library employs or plans to employ when ingesting born-digital records stored on legacy media. N=19

Current Strategy

A documented risk in our holdings is the presence of legacy media scattered throughout the textual holdings. We need a systematic means of accessing the data in these media to determine if the contents should be preserved as records in our holdings.

As new collections come in with digital content, we copy them to a server. Have not systematically gone back to find digital content in legacy collections, so those are being stored on legacy media.

Converting legacy files to modern file types.

For the most part, we are committed to access and will migrate files to the latest usable format to provide access to the content. We realize that we sacrifice the original look and feel of the files, as well as their functionality, but it is an acceptable loss given our main users for this content and the nature of our content so far (mostly word processing files). For example, when we have encountered legacy files on floppy disks, we have converted the files to the PDF-A format and made them accessible online. For materials with copyright or privacy concerns, these are available in a Virtual Reading Room, so just like our physical reading room in Special Collections, researchers must fill out an application form and agree to our terms before entering the Virtual Reading Room online to access the content.

One strategy: Using earlier (Mac) models to open older files, movies, etc. Holding some old software OS9, earlier versions of iMovie.

Remove data from media ourselves when we are able.

Storage on servers in a "Dark Archives."

Transferring content from legacy media to server storage.

We are selectively copying born-digital files (mostly photography) to our servers for backup. We hope these in turn will be moved over to the digital repository for digital preservation actions.

We have a sophisticated digital video encoding platform that enables us to retrieve video and audio data from a range of legacy formats (VHS, Digi Beta, etc.) For formats we cannot manage, such as reel-to-reel tapes, we occasionally outsource to a commercial vendor. Whether or not we retain the legacy media depends on whether it is accepted as a collection in our Special Collections/University Archives Department. If not, we return the legacy material to the collection owner. We are increasingly receiving research data in commercial formats such as Excel. Our current strategy is to document the version and test sample data with new software versions for backward compatibility and to store the data whenever possible in an alternative, less proprietary format. Currently, Excel and other database formats are also stored as CSV. We are looking at the DDI data format and other XML solutions as another non-proprietary standard. We are more interested in finding non-proprietary standards that retain the information content than in emulation or encapsulation of legacy data. Our biggest issues are research data formats proprietary to a specific data analysis tool, such as the FASTA format for gene sequencing, since we do not currently have an acceptable non-proprietary format for such data.

We have some legacy hardware but have no intention of building a true "collection." We use it to retrieve and transfer content from legacy media when possible. When we encounter a format we do not have hardware for, we turn to neighbor institutions for assistance; when this fails, we consider the likely value of the content on the legacy media. If it is not high, we will generally store the hardware as is. If it is high enough, we would consider outsourcing, dependent on cost and availability of funds.

We migrate digital media into a "digital archives" sever area that replicated our intellectual department divisions.

With some legacy media we can have our IT staff transfer the data, but I would not consider this outsourcing.

Planned Strategy

Migrate materials to newer media.

Plan to transfer data when resources are available.

Transfer to server storage (or, for example., repository).

Unknown. The planning process is just beginning.

We also have a strategy to investigate 3rd party vendors and their abilities for normalizing content for ingest.

We feel that the legacy hardware/software requirements for the digital content in our current "hybrid" collections are modest and can be addressed with local equipment. We have also purchased the "FRED" forensics system that will boost our capacity. We anticipate there will be some types of legacy media where we will need to use external vendors for content retrieval.

We plan to transfer records stored on legacy media to server storage and into the library's digital preservation repository. Two units are developing a collection of legacy hardware. One is outsourcing the process of retrieving the data.

Other Comments

We are currently pursuing a mix of 1, 2, and 4 but are interested in the fifth option and keeping track of collaborative efforts in the field.

We do not currently outsource the retrieval of born-digital materials from legacy media nor do we actively plan to, however, that doesn't rule out the possibility of doing so.

While we do not currently outsource the process of retrieving the data from legacy media, we have utilized this strategy in the past.

ETHICAL/APPRAISAL ISSUES

11. Below are ethical/appraisal issues that may be encountered while managing born-digital materials. Please indicate which of these issues are addressed by your library's ingest policies or procedures. Check all that apply. N=42

Whether to retain (under restriction) or destroy personally identifiable information (PII)	30	71%
Whether to preserve e-books, software, digital music, and other copyrighted content	20	48%
Whether to make files available for research use without having screened them for PII on the file level	18	43%
Whether to retain or destroy file fragments and deleted content in the absence of explicit guidance in the donor agreement	15	36%
Whether to preserve log files, preferences, browser caches and other types of ambient data in the absence of explicit guidance in the donor agreement	5	12%
Other issue	16	38%

Please describe the other issue. N=16

PII and Restricted Data

All special collections materials have personally identifiable information (PII). This may be different than sensitive information, which may be protected when PII cannot be.

Personal email, bills, contents of individual artists that we encountered.

Some records held by the Medical Center Archives contain Protected Health Information (PHI) and are covered under the HIPAA Privacy Rule.

There is an institute PII policy, but to my knowledge not a more formally written policy to specifically addresses managing archival materials.

We have policies that deal with copyrighted material regardless of medium and we have institutional policies that deal with PII and restricted data, but nothing that specifically applies to the collection of digital materials within the libraries.

While we have policies on privacy, we will need to develop more granular procedures for dealing with born-digital records that are in alignment with those policies.

Other issues

A variety of procedures are in place. Some address some of these issues; some do not.

All of these are covered by policies and procedures from across the Libraries, not just within the digital ingest group. As complex issues and concerns, these are not within any single policy and are instead supported by many policies and procedures.

Development of policy underway; donors have applied access restrictions.

Our holdings are unprocessed Presidential records that require access review and the completion of a notification process defined by Executive Order 13489 prior to public access to any of the records.

Related to appraisal, we are also trying to address whether or not this material is or will be deposited with another institution. Since the donor or depositor does not have to hand anything physically over to us, and even if they do, since they may easily make and retain a copy, we are concerned that we may be spending time and resources on material that is not unique and that the donor may wish to deposit in multiple institutions creating a redundant work to process the material in multiple places.

We are currently developing policies and procedures in this area in conjunction with the acquisition of our first major born-digital organizational archive.

We are in the process of creating policies for institutional records. These issues will primarily be addressed through file plans and retention schedules. We have not addressed these issues for personal materials.

We are just beginning to discuss these issues.

We have the ability to capture and maintain rights metadata, the IRB policies for specific research data. We can also control access to parts of a research data collection that need to be preserved but not made available for privacy or copyright issues. We hope to implement a dark archive in the coming year, but currently we will preserve born-digital resources that are in a fragile format (such as superseded video file formats).

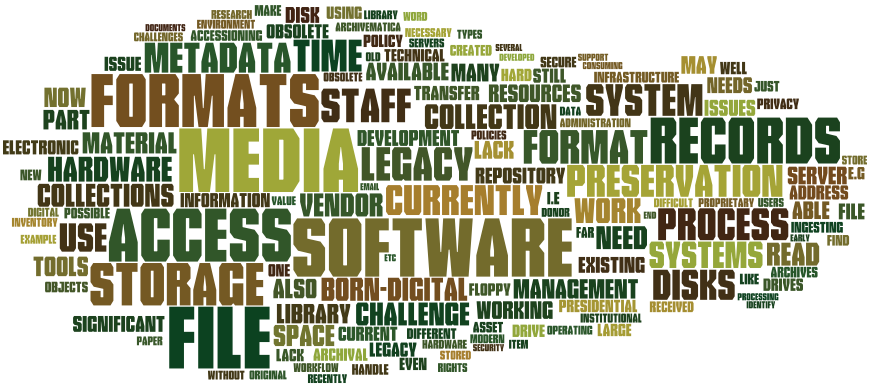
We would follow existing guidelines from the print world, I expect. Hasn't come up yet.

12. Does your library have a written policy that addresses your PII practice? N=59

Yes	12	20%
No	47	80%

INGEST CHALLENGES

13. Please briefly describe up to three challenge(s) your library has faced in ingesting born-digital materials (e.g., file format, hardware, software, privacy or security issues, etc.) and how the library has addressed that challenge. N=60



Ingest Challenges Word Cloud

A significant challenge has to do with legacy media. The oldest format we have so far identified are 8" floppy disks in a WANG format (i.e., not a modern PC format). We also have identified a number of 5.25" and 3.5" floppy disks, as well as CD, DVD, and hard drive formats. We suspect that there also may be data tape formats in the stacks as well (an inventory project is currently underway), and have been in talks with a donor with data on IOMEGA jaz and zip disks. We have been able to acquire media to read the 5.25" and 3.5" floppy formats, but not the 8" formats. We have also acquired a forensic imaging device that can handle a number of data connector types such as SCSI, SATA, and IDE. When we are able to physically read a disk, some of these devices have controller cards or software that work only on specific operating systems, making it difficult sometimes to physically read disks in an efficient workflow (because some disks in a collection are read with one machine and one software, producing one kind of output file, while others need to be read on another machine using a different software and output file). For those pieces of media that we currently do not have hardware for, we have to evaluate the costs of acquiring the hardware against the value. For example, we could possibly find drives to read the 8" disks, but we are not guaranteed to find a controller card to make that drive compatible with a modern PC anyway. In this circumstance, we have therefore investigated the prices of having a vendor image the disks for us, but this also requires a cost-benefit analysis. Each disk contains less than 80 KB of data and the price just to image the disk is around $50 each. In addition, we need to ship the disks and we run the risk of having them lost or damaged in transport. As well, we would need to have some sort of confidentiality agreement with the vendor regarding the privacy of the data because we have no real idea of what is on the disk. After all of that, we could then send these disks to the vendor and find that they are unreadable anyway – there is no way to tell if a disk is readable until you attempt to read it. Finally, the issue of the scale of this legacy media is a challenge. We estimate currently that we hold less than 3,000 disks, but the time necessary just to load and transfer data from those disks is considerable. This is not even including the time it would take to process the materials as part of a collection, but simply to transfer it from the media where it is at a higher risk of corruption, to network storage. A second challenge related to the first, but separate, is the ability to actually read the data that is on the media. As described above, we have some material that is so old that it is not even readable by a modern computing system. Other data is not as old, but still

presents significant problems with readability because the data is stored in file formats that are no longer compatible with modern operating systems or for which we simply don't have software to read. An example is architectural drawings created by a CAD software in the early 1990s. In this case we are able to work with our School of Architecture to locate some software to read these programs, but it does mean that we need to keep this software viable, which in many cases means running older operating systems or alternative operating systems to what we currently use on the forensic imaging hardware (which is primarily Windows). One of the pieces of software we have purchased is Forensic Tool Kit, which can identify and "read" thousands of file formats. However, these formats are primarily those that would be most commonly seen in criminal investigations, since that is what the software is designed for. So, things like CAD software from the early 90s are not included in their list of recognized formats. We have not seriously discussed trying to emulate any software or operating systems at this time, although we have watched with interest other projects that have done so. We do not view emulation as a viable approach at this time since our collections are so diverse and we do not have the type of technology staff in the library to really do this work efficiently. It would simply be impossible to have the resources available to emulate each and every program we are likely to encounter and to keep those emulations running in current environments. While there are some things we are likely to see a lot of (Microsoft Word documents, for example) we also feel that it is not worth the effort at this time to create an emulated environment when a migrated format (a PDF in this case) would be adequate. This is not to say that in the future emulation may not be attempted in special circumstances. A third very significant challenge is related to the lack of available tools for doing archival work with born-digital collections, as well as infrastructure in terms of repository and preservation networks that can meets the needs of access, management and preservation. There are several open-source and commercial products that can do pieces of the workflow, but as they are not designed to work together there are inefficiencies in stringing these workflows together. As an example, we use the Forensic Tool Kit software to extract some basic technical metadata, identify duplicate materials, and those that might contain predictable sensitive information such as SSNs or credit card numbers. The output of FTK, however, is some proprietary XML and a PDF report. We then use Archivematica to further extract technical details and establish a provenance through the creation of PREMIS metadata. We would then record information about a duplicate removed from the accession in Archivematica, but ingesting the duplicate file and then removing it manually per the FTK report. Finally, the PREMIS metadata record that Archivematica creates in nested inside a METS record for the entire accession. Our current storage network however, wants only the individual PREMIS records for each file, rather than the combined METS, so more work needs to be done to transfer the file between these two tools. Once the material goes through this network of tools, we still need to work on our repository and other digital asset management and discovery systems in order to suit the needs of this material which differs in many ways from the needs of other digital materials we store and manage such as e-books and –journals and digitized resources. This infrastructure needs to handle the preservation, management, access, and discovery of these materials. We are watching with interest the developments of open-source tools created by the archival community such as Archivematica, bitCurator, ArchivesSpace, and Curator's Workbench as well as potentially doing some work on the further development of Hypatia.

Adequate digital infrastructure to securely store and describe born-digital content. Adding these responsibilities onto existing staff: training, workload. No formal records management policy at the university.

Appropriately secure storage. Staffing resources. Policies and workflow development.

Copying/reformatting from old redundant file formats. Network latency and storage; lack of server space. Lack of software to support integrity of file reformatting.

Copyright: all our metadata contains a copyright statement for our digital object. Other options we can apply are banding and watermarking to objects. We include the copyright holder when it isn't our institution and we know who that is, but this becomes a challenge when unknown. In some instances, we have put up digitized objects, asking for input from our patrons for ownership. Fixity: we don't currently have a systematic way of guaranteeing fixity! We

are actively working on a preservation plan that will address this issue. Authentication: we don't currently have a mechanism to authenticate born-digital objects – we "trust" the source and ingest. We are hoping to make this part of our Digitization Preservation Policy, which is currently in development.

Developing policies and procedures relating to the acquisition and ingest of born-digital content: the Digital Archivist has recently completed a research leave where he has drafted a digital preservation policy that could apply to born-digital materials. Developing an open-source digital asset management system: the ingest process for our digital asset management system has been unreliable in its early stages of development. The Libraries has dedicated an IT person to this system and has hired a vendor to further development of the system, particularly regarding its stability. Creating an inventory of born-digital material on legacy media: the Digital Archivist will soon be compiling such an inventory based on existing finding aids.

Developing secure hardware infrastructure to protect PII collected and retained; have worked closely with the campus IT security office. Securing secure, backed-up server space for dark archive. Planning access strategy for restricted content.

Digital storage space. We have recently conducted an inventory of all of our special collection digital assets (not just born-digital). This will be used to more effectively plan our storage needs—the amount and types of storage. Sustainability of digital library and preservation platform. We haven't yet adequately addressed this issue.

File format is an enormous challenge. We are receiving research data proprietary to specific data collection and analysis tools, such as the SURF surface mapping data produced by the software MountainsMap. Another is the gene sequencing data, FASTA, produced by the SOLiD gene sequencing system. We don't have non-proprietary formats in which to store this data and we don't know enough about persistence and backward compatibility for the tools. Our researchers are skilled at using the tools and interpreting the data but aren't able to answer our questions about persistence and longevity for the data. Thus far, our only strategy is to document the instruments that created the data, document as much as we know about the data (which is often in multiple files) and to bring this issue up in every research data gathering and suggest that conversations with these instrument providers are needed. File size is another challenge. Large files take a very long time to process and can make born-digital files difficult to manipulate in the repository and for end users to download. We currently bundle large files into zip files for downloading but need an effective background methodology for ingest.

File format on legacy tape drives from punch card data that has Census/private information for different nations. Need for old hardware on site for conversions and ingest with immediate time demands. Scaling up for the demand.

File formats: i.e., Word 1.0 documents. Hardware: i.e., receipt of records on 5 1/4" or 3 1/2" discs; no computers that will read such discs. Uncertainty about the authenticity of the records we have received. Do we have the only copy or are there multiple copies/versions available elsewhere?

Hardware and software. We don't always have the hardware and/or software to access legacy file formats, and don't know how to access files without changing their metadata. We try to collect obsolete hardware when possible, and sometimes outsource accessing these legacy files. Selection of file formats for streaming media; we are currently working on this with library IT staff. We face challenges trying to educate the university community about giving us their born-digital files, and lack confidence that we can preserve it and make it accessible because of lack of resources and internal technical expertise. We are working on outreach to university offices, and working on developing necessary skills for archiving born-digital content.

Hardware; lack of secure storage and backup. We are attempting to implement now, working with university IT. Privacy/security. We hope to develop written policies.

Images received in digital format but named idiosyncratically by the photographer. In order for these files to be used in a local digital environment it is necessary to provide meaningful file names in relation to existing or new local directories.

A procedure using a combination of Adobe Photoshop and Adobe Bridge was developed locally to batch process files to accomplish this task. Sensitive data: we have yet to work out issues surrounding born-digital institutional records with restricted access, e.g., promotion & tenure files, President's Office files, etc. An organization uses an online service to process applications that in the past had been delivered in paper format. Acquiring the records in a format that is useable by the archive may require a contract of some sort with the vendor. This remains to be resolved.

In 2010 the library acquired a collection of nearly 50 floppy discs and a number of CDs; most were unlabeled (or labeled unhelpfully), meaning that we had to view each one and try to deduce at least minimal information so we could describe the contents. However, the most challenging item was a hard drive, carefully wrapped, with a label reading "The contents of this drive can only be accessed at the original computer from the New York Times. If installed at any other computer, you may damage the contents and you may format (wipe out) the drive." We have no idea quite how to approach this so have simply left it alone as is!

Inability to access content saved on obsolete media or in obsolete programs. Lack of secure, redundant, geographically distributed, and reliable preservation storage systems. Lack of system for managing and providing access to born-digital materials that will allow for restricting some content for a period of time and will also help automate some processes like generating checksums, virus checking, extraction of technical metadata from file headers, etc.

Ingestion of compound/complex objects (i.e., objects made of many types of materials at once). We use Google Spreadsheets to compile metadata and file locations, but a solution like BagIt is likely to be more effective. Presentation of complex objects. Determining how to show a user an object consisting of many disparate parts (e.g., a video with a transcript, screenshots, and an associated web page). This is usually considered a prerequisite to ingestion, since an object is only considered accessible if it can be usefully retrieved. We still address this question on an ad hoc basis. Providing granular security options for all content. The technology required to provide very granular control over rights and permissions makes it difficult to build services for ingesting and reusing repository content. Few repository systems (we use Fedora) have a fully developed solution in this regard, so we use our own solution based on the university's Shibboleth identity system.

Lack of a standard set of best practice guidelines for dealing with original context (e.g., file system hierarchy) of born-digital files when ingesting. Lack of a policy on file format normalization, and identification of what a "record copy" means in the born-digital context. Fear and misunderstanding of the nature of born-digital material.

Lack of software and/or hardware to read files and physical media: We rely on library and college IT departments to access file content, and we acquire legacy hardware when possible. Lack of server space to use for transfer of records from digital media: We recently acquired server space hosted by the university's IT department for use in backing up digital media. Maintaining privacy and security of confidential records; complying with university policy as well as federal and state laws governing privacy: We have policies governing access to confidential records, but procedures specific to born-digital materials are still being developed.

Legacy File Format Normalization: We have a collection that includes over 25 different file extensions, mostly text-based documents, many of which were unrecognized and/or created significant artifacts or "garbage" when rendered in modern programs. A lot of these files were created on the now defunct and unsupported Nota Bene annotation/bibliography software. We used a conversion tool called FileMerlin to convert as many of the troubling files as we could and a Windows Command Line script utilizing Microsoft Word to convert Wordperfect and other Legacy File formats that Word would recognize. After a significant amount of manual and automated work, we increased the number of legible files in the collection from around 40% to around 95%. Legacy Media recovery: Like many institutions, we have many "hybrid" collections that include legacy media such as 3.5"/5 1/4" floppies, hard drives, CD/DVD, even whole computing environments. We are building a Legacy Archival Media Migration Platform (LAMMP) and an accompanying manual as an environment and a workflow for capturing images of these media and generating metadata and capturing

contents where possible. We have finished developing and testing this process and are ready to image our first batch of 3.5" floppies, followed shortly by 5 1/4" floppies and hard drive after we acquire the hardware (drives, drivers and write blockers).

Legacy software: needed legacy equipment to access and transfer files. Donated mixed material collections: donor may not own rights to all of collection that was contributed. Images in dissertations that might have fair use rights but not necessarily general dissemination rights: how to deal with this.

Limited staff comfortable with ingest. Although we have an ingest process that has now been formalized and undertaken with more than 50 accessions, we still only have a couple of staff members who possess the sufficient technical skills and understanding of digital records issues to undertake even the rudimentary steps in the accessioning process. This leads to resource constraint issues as more and more digital records on media are being taken in, even if they are not actively collected. To grow this program more, we need more, and lower-level staff to undertake much of the accessioning process, as they currently do with paper. Minimal description practices don't match ingest process. We are following and forensic model of accessions where we are creating forensic images of storage media during accessioning and setting those images aside for further processing. However, the current model for archival accessioning on paper is to undertake minimal arrangement and description during the accessioning process, thereby eliminating a backlog requiring future processing. Hardware and software ingest lab development was time consuming and difficult. Although we have now built up a significant shared lab to enable the ingest of born-digital records from many different types of storage media, the process of building such a lab took several years, expertise, and funding. Each new collection seems to bring new technical issues that must be dealt with.

Major issue is technological — especially how to receive content from private donors. Still being worked on.

Media obsolescence/failure. This includes outmoded storage systems like 5.25" floppy disks and zip disks. Even if we have hardware to accommodate them, we sometimes find that the content is corrupted or otherwise inaccessible. We have a small collection of old drives and other resources nearby; after that we consider outsourcing but will often store as is or even deaccession, depending on resources and anticipated value of the content. Software obsolescence: sometimes it isn't even obsolete, it's just got a small market share, like AskSam. So far, we have been able to find programs to access and migrate/normalize this content. File formats: we have received proprietary camcorder files, for example, which we had difficulty assessing the value of. Upon further investigation, these were found to be metadata files and thumbnails. We determined in the end that we would keep them.

Met with outgoing dean and transferred email account to library servers once he left the position. Outlook PSTs are highly proprietary. Transferred deceased faculty member's email account to library servers. Mac to Windows migration was very time consuming. Email account is Eudora and no easy way to convert emails to less proprietary format. Transferred digitized president's office correspondence from CDs to library servers. Transfer process took hours.

Obsolete file formats. Readability of legacy media. Lack of identifying information accompanying legacy media (unlabeled, no contextual information).

Obsolete media storage. To date we have been able to outsource this to a vendor. Lack of any repository to store or manage personal materials donated. We've taken in a few batches of material and have stored them with only a promise of byte stream recovery and have temporarily turned other material away.

Obsolete media, file systems, and file-formats; e.g. 8" floppy disks, FAT variant disk formats, and WordStar files (existing converters did not work). Data loss from media corruption. Managing the politics surrounding SEI/PII. Some disks have content the donor did not expect to be there, was private, and outside of our collecting scope. Some capture mechanisms are poor or incomplete compared to the original versions; e.g., social media and enterprise systems data.

Obsolete or deteriorating storage media: floppy and optical disks. We are in the process of transferring materials received on such media to networked storage, to facilitate bit-level preservation. Obsolete or unknown file formats. We currently rely on available open-source file identification tools, the file-conversion features of desktop applications, such as MS Word, and the expertise of contract staff familiar with the history of common office work applications. Metadata capture and management. Metadata for born-digital special collections materials is currently managed using the Special Collections databases for accessioning and archival description information. We plan to transition in the near future to a digital repository application with more sophisticated metadata management functionality.

One of the biggest challenges we have faced with our collections is how to satisfactorily handle security and privacy concerns of our donors. Because it is still early days with born-digital personal collections, we are approaching this problem by proposing a process that ensures donor confidence, reviewing outcomes then suggesting other approaches that enable more sustainable practices while also addressing donor concerns. We continue to struggle with identifying archivist-friendly tools to use for ingest and processing. All we can do for now is follow development of tools and best practices in the field. Securing dedicated staffing for digital archives work continues to be a challenge. We now have one dedicated staff member, the promise of another dedicated archivist, and support from other library divisions. Advocacy with senior management about the needs and importance of digital archives has been our only approach.

Opening Legacy formats. Privacy/Security issues of PII. Workflow for ingest.

Organization: Research Data and Institutional Archives come on unorganized file systems. Files need to be restructured into standard, flattened directories representing collections of items. This requires significant analysis and scripting. Metadata: ETDs come from the vendor with transformable metadata. However, metadata is usually non-existent in Research Data and Institutional Archives. Sometimes it can be derived from full text from documents. In the case of audio, video and images, it must be entered or derived from external spreadsheets. Disk Space: Archives are guaranteed preservation only if stored on enterprise data storage. Redundant, highly available enterprise disk is still costly. Traditional administrative systems use relatively small amounts of storage so the infrastructure must ramp up an order of magnitude. An entire integrated library system runs on less than half a terabyte while Research Data collections often utilize 1–10 terabytes each. This creates issues around scheduling and funding disk space acquisition.

Organizations of born-digital material. Have created separate master list that contains organization data. Archiving data as brought in in appropriate format. Standardizing metadata and quality control.

"Preservation environment" vs. "Repository." "Repository" is currently under development. Unwritten vs. written policy (Policy is still underdevelopment). "In permanent develop." Most ingest software is in an alpha or beta release, with long-term roadmaps for future development.

Pressure from donors and partner institutions that want us to be able to handle all existing file formats. We're just beginning to grapple with questions of workflow, appraisal, and how to make the files available for research use. Metadata: how much and who creates it? So far we've received a lot of assistance from our technical services unit for materials being ingested in our IR, but we can't expect them to take on that burden for born-digital archival materials, emails, etc.

Privacy: partially addressed through limited access by staff and warnings that material is restricted.

Readability of legacy media: some disks that were accessioned in the past are now unreadable. We currently do not have a strategy to address this. Appraisal: some legacy media acquired in the past was accessioned as part of a larger collection without thought to whether the disk content has sufficient research value to warrant preservation. We now need to decide whether or not to reappraise this material. File format: legacy media has a variety of file formats, many of which are no longer in use. We are piloting the Archivematica preservation system, which will normalize some formats into an access and/or preservation standard.

Redundant storage. Obsolescence. File/data management.

Right now, the libraries do not really have a plan and they are all ingesting born-digital files in different ways. Mostly, materials are kept on external hard drives or in their original legacy media.

Server space: inadequate space for our current digital projects and lack of understanding from administration that there is a need for special collection to have their own serve for processing digital collection. Digital preservation: would like to be part of a LOCKSS system. Currently we are saving our digital collection on main library server as well as the university's server. Next step LOCKSS. Hardware: acquiring legacy computers, working with other departments on campus to identify and locate existing hardware.

Stable storage: large enough; offers growth; limits access. Have just implemented new Isilon mass storage utility. Born-digital mystery files. Purchasing legacy hardware; need forensics software. Workflow to ensure preservation master and deliver use or access copies. With installation of Isilon, parts in place, more discussions underway.

Staffing and time required. Funding. Technology resources.

Staffing: lack of staff positions to address preservation of digital and born-digital content. Challenge addressed in part by re-defining a vacant archivist position as an "e-Archivist" position. We also write additional staff positions into grants wherever feasible (currently we have 4 FTE on grant-funded appointments). Also, existing staff have had new responsibilities added to their job descriptions to support digital preservation efforts. Legacy hardware/software: first tests of born-digital content found in mostly paper archival collections had a low success rate of content acquisition. Challenge will be addressed in part by identification of legacy devices and software by our IT group and in part by our purchase of the "FRED" digital forensics package. We have not yet done a detailed inventory to identify and document types of born-digital content in our collections. Storage: our current archival storage infrastructure was scaled to accommodate our analog to digital digitization program. We projected needing 70TB (replicated 3X) to serve needs through the end of FY2013. Now newer born-digital preservation and access projects will take us far beyond 70TB. Our response has been to begin planning now for a significant increase (to 250TB) at the beginning of FY2013. All funding will be reallocated from existing library budgets.

Storage space: library IT has bought new servers, and the library is collaborating with the campus Center for Advanced Research Computing for additional server space. Permissions (privacy): it will continue to be a challenge to maintain appropriate access and privacy permissions. We are working with various systems (DSpace, CONTENTdm) to explore possibilities of restricting access to specific individuals. Findability: data made available through online servers needs to be findable. We are adding metadata records to our institutional repository that describe datasets available online.

Storage that is secure, backed-up properly with sufficient room for growth. Appropriate workflows to ensure the accurate ingest of born-digital materials. Playback equipment that can extract content safely. Allocation of staff time to focus on issues, develop policy and conduct test pilots to ensure a more proactive process.

Technological support: have to find resources (storage) where we can.

The biggest challenge that our library has faced in obtaining the hardware and software necessary to ingest and manage born-digital materials. We are currently beginning a pilot project using Rosetta and hope that this will enable us to better handle born-digital materials. The second major challenge is the issue of personally identifiable information (PII) in born-digital collections and what to do with it. We currently have no policies for dealing with PII but hope to devise some as part of our Rosetta pilot project. The third major challenge is training our curatorial staff how to deal with born-digital materials--this includes ingesting born-digital materials. We have been working actively with the Conference of Inter-Mountain Archivists to bring several of SAA's Digital Archives Specialist (DAS) courses to our region and are strongly encouraging our curators to participate in these workshops.

These challenges from Presidential Libraries are representative of the challenges other parts of NARA also experience. Volume of data to be ingested in as short a time frame as possible: We receive the vast majority of our electronic records in large transfers at the end of a Presidential administration. Because of our need to provide asset-level access to electronic records as soon as the records are in our legal custody we need to ingest these large volumes in as short a time frame as is possible. In our last large transfer we worked with the records creators and with our system vendor to devise a means of transfer that employed storage area networks (SANs) to move large volumes (tens of terabytes) of data copied from the creator's data center to the data center for our Electronic Records Archives (ERA). Four physical shipments of data stored on SANs over the course of several months moved more than 70TB of data from the source data center to our data center, where the files could be staged for ingest and then moved into our system environment. File-level access control policy: Our system users are located across the country. All users fill the same role in the system, but users should have access to only subsets of the electronic records maintained in the system (Presidential records from one administration versus Vice Presidential records from another administration, for instance). To maintain asset-level access control (among other needs) we established asset catalog entries (ACEs) that were assigned to each asset upon ingest. These ACEs (xml files) include elements that define each asset by a Presidential administration and by a records status (Presidential, Vice Presidential, or Federal). When users log in to Executive Office of the President instance of the Electronic Records Archives (EOP ERA) the system is able to compare the rights of the user to the characteristics of assets to determine if the user can have access to the files. Need to make electronic message files accessible: The storage architecture deployed in EOP ERA makes hundreds of formats available for indexing, including .eml files for emails. One set of electronic messages planned for transfer to us during the last transition (more than 20 million files) was stored in a journal format that maintained the messages as text files. Because we wanted to access the messages as emails (i.e., using parametric searches of email fields – To, From, Date, etc.) our vendor (Lockheed Martin) developed a script that transformed the text files into discrete .eml files that could be ingested into EOP ERA and managed as email files. As part of this transformation process the vendor used sample data to inform a discussion with our archivists on the fields we wanted to maintain in the .eml target files. As part of testing we were able to assure ourselves that the content of the messages came through the transformation intact, including any files attached to the original message files.

There is no Digital Asset Management System (DAMS) in place to ingest born-digital material. System wide initiatives would address this problem. The necessary hardware to transfer born-digital material from legacy media is not available at our repository. A few pieces of legacy hardware have been purchased. Staff expertise to deal with ingesting born-digital materials is limited. This has not yet been addressed.

Time: Reformatting legacy media, and arranging and describing born-digital content, are time-consuming activities. The volume of data that can be found within a single item such as a hard drive can be staggering. Migrating content from legacy media is also time consuming as there is little automation/batch handling of these materials. We are investigating ways in which to reduce time spent on individual items. Migrating unidentified content: With unidentified content on an obsolete media format it's difficult to determine whether the content is a reformatting priority without accessing the material. If we do not have the equipment in-house for the obsolete media format the item requires access by a vendor. Sending an item out to a vendor is expensive and may not be the best use of our resources. At this point, we are investigating ways to address this issue without overuse of resources. Software licensing: Due to stringent state regulations on software purchasing and needing obsolete software titles to access files that may be generations removed from current software (or without a contemporary equivalent) acquiring appropriate software necessary for file migration is a challenge. We are looking into software titles that can bridge generations; that is, software that can open older files and convert them to a newer generation that can be accessed with current software. We are also examining software designed to open obsolete file formats such as Quick View Pro.

Training of existing staff and addition of trained staff to handle the quantity of incoming digital materials expeditiously. Better administrative interface and workflows for staff members ingesting born-digital content. Appraisal of an increasing volume of born-digital materials efficiently.

Unknown file formats. Inadequate software for specialized file formats (e.g., CAD files).

User contributed file formats: some of the content is not is a standard format. Talk to potential donors about contributing content that confirm to open standards. File size: one of the platform that we use is hosted DSpace. If files are too large to upload we work with the vendor to load materials. Restricted items: we try to restrict the materials so that they are available to certain communities.

Variations in file formats, packaging, naming schemes. Applications needed to access content. Lack of clear preservation policies and procedures.

Visible vs. dark archiving. Larger institutional inertia on issue of electronic records management.

Volume of materials, how to appraise. Quality of data, e.g., image files that have low resolution. Not address how to provide access to digital materials when associated with analog collections.

We are managing somewhere between 50,000 and 100,000 digital files on media and server space. We are attempting to copy files from media to server to ensure backup. We have only recently been given permission to load materials to the digital repository, but we have received no additional staff to produce metadata at the item level. We have four pilot projects in progress using paraprofessional staff and interns for metadata production. We want to continue collecting certain basic university publications (i.e., course catalogs) that are formerly paper and now either database driven or web publications. We are negotiating workflows and agreements with producing offices and vendors to produce a continuous online backfile of certain critical titles.

We are working out issues relating to born-digital materials and have not encountered significant challenges with what we have done so far, postponing the more problematic aspects until we get there.

We have born-digital materials on CD and DVD for which there is no server space or metadata provided by the creator of the materials. We address this through a redundant array of external hard drives and back up that is merely a stopgap solution to the problem. We have no expressed authority or access to most born electronic records in other systems such as Banner, so there is no way to review such records for historical value. An ad hoc records advisory committee recently approached Administration requesting creation of an electronic records committee with oversight authority to address these issues campus-wide.

Whether or not the quality of the born-digital is up to par with our institutional benchmarks and guidelines for digital media. In some cases, re-capture is not possible. Discussion with our working group will then include whether a poor copy will be included in the digital library or not. Dealing with file formats that may or may not be compatible (or able to be migrated) with current guidelines of institutional practice. We will test the file to see if a comparable format is acceptable or if data is lost during this process. Sometimes, this will allow our group to explore different presentation tools for other file formats, or we have the option of storing the file only (no automatic presentation tool).

STORAGE POLICIES AND PROCEDURES

14. Please briefly describe who is responsible for each of the following storage activities/functions (e.g., special collections/archives staff, library IT staff, parent organization IT staff, etc.). N=63

Selecting Storage Solutions

A combination of Special Collections/Archives and library IT staff	18	29%
Library IT staff	14	22%
A combination of library IT and campus IT staff	6	10%
Campus IT staff	3	5%
A combination of Special Collections/Archives, library IT, and campus IT staff	3	5%
Other	19	30%

Ad hoc committees led by central library IT

Archives, digital curation leadership, and library network/system administration

Digital Services and Shared Collections Department and Digital Development and Web Services Unit, in conjunction with institution-wide IT

Digital Strategies committee (drawing from library IT, Special Collections, and other units within library)

Electronic Records Archives Program Management Office (within Information Services) and Preservation Staff (within Research Services)

For the repository, the Director of Integrated Information Systems works with a team, including network administrators and the digital data curator. The Special Collections/University Archives staff will lead a library (possibly university) team to select an electronic records management system for the university.

Institutional records: a team consisting of the records manager, college archivist, archives staff and institutional IT. Personal materials: college archivist, special collections technology coordinator, manuscripts supervisor.

IT staff at the Southwest Collections/Special Collections Library, Digi Resources Library Unit, and the University Library

Library & parent IT staff/consortia

Library administration (once Library Systems recommends)

Library Information Technology Office and Library Digital Programs Division

Library IT and the Carolina Digital Repository

Library IT and the Office of the CIO

Library IT staff, in consultation with Special Collections and Preservation/Digital Initiatives

Library IT/Digital Initiatives

San Diego Supercomputer Center (SDSC)

Selecting the technology for the institution is handled by our IT Department. Selecting the appropriate tools from the libraries available resources is handled by the curatorial and program management staff.

University Libraries Central Operations department, Digital Preservation Strategist

We have a Technical Architecture Council that works in concert with collection managers and IT staff to select.

Implementing and Maintaining Storage Infrastructure

Library IT staff	28	44%
A combination of library IT and campus IT staff	14	22%
A combination of Special Collections/Archives and library IT staff	5	8%
Campus IT staff	4	6%
A combination of Special Collections/Archives, library IT, and campus IT staff	1	2%
Other	11	18%

Archives, digital curation leadership, and library network/system administration

Central Operations department, Digital Preservation Strategist

Digital Services and Shared Collections Department and Digital Development and Web Services Unit, in conjunction with institution-wide IT

Electronic Records Archives Program Management Office (within Information Services) and Preservation Staff (within Research Services)

Institutional records: a team consisting of the records manager, college archivist, archives staff, and institutional IT. Personal materials: library IT.

IT and Scholarly Publishing and Data Management Team

IT staff at the Southwest Collections/Special Collections Library, Digi Resources Library Unit, and the University Library

Library IT and the Office of the CIO

Library IT staff, campus IT, California Digital Library Staff

San Diego Supercomputer Center (SDSC)

Special collections and preservation librarian

Managing Permissions/User Authentication.

Library IT staff	20	32%
A combination of library IT and campus IT staff	8	13%
A combination of Special Collections/Archives and library IT staff	6	10%
Special Collections/Archives staff	3	5%
Campus IT staff	2	3%
A combination of Special Collections/Archives, library IT, and campus IT staff	2	3%
Other	22	35%

A combination of special collections, library administration, library IT, and parent organization IT

Archival Staff \ Digital Initiatives Librarian

Campus/library network/system administration

Central Operations department

Digital Initiatives/collecting unit

Digital Services and Shared Collections Department and Digital Development and Web Services Unit, and Special and Area Studies Collections for permissions to physical materials as they are transferred to digital.

Electronic Records Archives Program Management Office (within Information Services) and Preservation Staff (within Research Services)

Implemented by Libraries IT staff; those with access to secure archival space must be designated by the director of a given unit or her designee.

Institutional records: institutional IT. Personal materials: library IT.

IT staff at the Southwest Collections/Special Collections Library Digi Resources Library Unit and the University Library

Libraries IT staff, Digital Archivist

Library & parent IT staff/consortia

Library Information Technology Office and Library Digital Programs Division

Library IT with input from collection staff

Library IT, curators, and the Office of the CIO

Library Research & learning support unit (where the institutional repository librarian is located)

Non IT digital repository managers and IT

Permissions: Library IT staff. User authentication: Campus IT staff.

San Diego Supercomputer Center (SDSC)

Shared responsibility between the IT Department and the system owners. May also be based on license or other agreements governing our content.

The library works with the Office of Information Technology to use LDAP, CAS and Shibboleth IdM for single sign on, but may also implement authentication microservices for specific projects. These are designed by the digital library architects and implemented by digital library programmers.

We're still working through these issues. We use the campus LDAP for our institutional repository. Other instances are managed by library IT and/or Digital Initiatives librarians.

Estimating Storage Needs

A combination of Special Collections/Archives and library IT staff	16	25%
Special Collections/Archives staff	11	18%
Library IT staff	9	14%
A combination of library IT and campus IT staff	1	2%
Campus IT staff	0	—
A combination of Special Collections/Archives, library IT, and campus IT staff	0	—
Other	26	41%

Archival Staff/ Digital Initiatives Librarian

Archives, digital curation leadership and library network/system administration

Central Operations department, Digital Library Services department

Collection staff

Digital Archivist, Libraries IT staff

Digital Collection Managers

Digital Initiative Librarians, Data Curation Librarians, Library IT

Digital Initiatives/Library IT

Digital Services and Shared Collections Department and Digital Development and Web Services Unit

Directors, Library/Archives staff, IT staff

Each collector/selector

Electronic Records Archives Program Management Office (within Information Services)

Institutional records: a team consisting of the records manager, college archivist, archives staff and institutional IT. Personal materials: library IT.

IT Department, based on input from the curatorial and program management staff.

IT staff at the Southwest Collections/Special Collections Library Digi Resources Library Unit and the University Library

Library & parent IT staff/consortia

Library Digital Programs Division

Library IT and curators

Library IT staff with Library Research & learning support unit

Library IT staff/collection curators/reformatted content producers (e.g., those migrating content from obsolete media into a modern format.)

Library Systems with input from Digital Library Services and Special Collections & University Archives

Research Data Curation staff

Special Collections to project what and how collections will grow, library IT to project what resources are needed and cost.

Special collections/archives staff, digital collections staff, library IT

Special Collections/preservation librarian

The Director of Integrated Information Systems and the Digital Data Curator

Budgeting Storage Usage

Library IT staff	22	36%
A combination of Special Collections/Archives and library IT staff	11	18%
A combination of library IT and campus IT staff	3	5%
Special Collections/Archives staff	2	3%
Campus IT staff	0	—
A combination of Special Collections/Archives, library IT, and campus IT staff	0	—
Other	22	36%

Administration

Archives, digital curation leadership and library network/system administration

Budgeting storage usage is determined at this time by Library Administration, Library IT staff, and Special Collections and Archives staff.

Central Operations department

Currently, the libraries are not allocating storage usage by project except in cases where grant funding has purchased specific storage amounts.

Digital collections

Digital Initiatives Librarians, Library IT

Digital Initiatives/Library IT

Digital Services and Shared Collections Department and Digital Development and Web Services Unit, in conjunction with Fiscal Services

Directors, Library/Archives staff, IT Staff

Electronic Records Archives Program Management Office (within Information Services) and Preservation Staff (within Research Services)

Head of Systems and Director of Administrative Services

Institutional records: a team consisting of the records manager, college archivist, archives staff and institutional IT. Personal materials: library IT.

Libraries IT staff, Digital Archivist

Library Administration, Library IT staff

Library Information Technology Office and Library Digital Programs Division

Library IT and curators

Library Technology Council (represents all stakeholders for technology issues)

N/A; based on pay for use.

Parent IT staff/consortia

Research Data Curation staff

Southwest Collections Administration and Library Technology Management System

Monitoring Storage Usage

Library IT staff	25	40%
A combination of Special Collections/Archives and library IT staff	10	16%
A combination of library IT and campus IT staff	5	8%
Special Collections/Archives staff	2	3%
Campus IT staff	0	—
A combination of Special Collections/Archives, library IT, and campus IT staff	0	—
Other	18	29%

Central Operations department

Digital Archivist, Libraries IT staff

Digital collections

Digital Initiatives Librarians, Library IT

Digital Initiatives/Library IT

Digital Services and Shared Collections Department and Digital Development and Web Services Unit

Electronic Records Archives Program Management Office (within Information Services) and Preservation Staff (within Research Services)

Institutional records: institutional IT. Personal materials: library IT.

IT staff at the Southwest Collections/Special Collections Library, Digi Resources Library Unit and the University Library

Library Digital Programs Division

Library IT staff monitor and advise Special Collections and Archives staff regarding storage.

Library IT staff with Library Research & learning support unit

Library Systems, Digital Library Services, Special Collections & University Archives

Not currently undertaken.

Parent IT staff/consortia

Research Data Curation staff

Special Collections/preservation librarian

The Director of Integrated Information Systems and the Digital Data Curator

Budgeting Storage Funding

Library administration	14	22%
Library IT staff	14	22%
A combination of Special Collections/Archives and library IT staff	6	10%
Special Collections/Archives staff	3	5%
A combination of library administration and library IT staff	2	3%
A combination of library IT and campus IT staff	2	3%
A combination of library administration, Special Collections/Archives, and library IT	2	3%
A combination of library administration and campus IT staff	1	2%
Other	18	29%

Associate University Librarian for Digital Library Systems

Central Operations department

Digital Initiatives/Library IT

Digital Services and Shared Collections Department and Digital Development and Web Services Unit, in conjunction with Fiscal Services

Digital Strategies Committee

Directors

Each collector/selector is responsible, although few actually undertake this task.

Head of Systems and Director of Administrative Services

Institutional records: institutional IT. Personal materials: library IT.

IT staff and Scholarly Publishing and Data Management Team

Library Information Technology Office and Libraries Administrative Services

Library IT staff, and Associate University Librarian for Digital & Discovery Services

Electronic Records Archives Program Management Office (within Information Services)

Parent IT staff/consortia

Requested an annual basis—IT and library staff.

Research Data Curation staff

Shared by library departments.

Southwest Collections Administration and Library Technology Management System

Other Storage Activity/Function

Arts Library work stored on local HD or external HD until pushed to I&D.

Digital preservation activities are carried out by special collections/archives staff working with Libraries IT.

Planning storage architecture: Library IT with Carolina Digital Repository

Special Collections and Archives are responsible for managing physical storage of legacy media containing born electronic materials.

Tape storage: Archives staff

We have a couple of strategies for ensuring files that are persistent and authentic, including multiple online, nearline, and offline copies and regular signature verification for each preserved file.

15. Please indicate which of the following storage solutions your library uses for ingest, processing, access, back up, and long-term "dark" storage. Check all that apply. N=63

	Ingest	Processing	Access	Back up	Storage	N
External Media Library (e.g., CD/DVDs, tapes, loose drives)	41	18	27	31	16	59
IT-supported Network File System	35	43	43	44	26	58
Local/Attached storage (e.g., internal drive, external drive or other local storage device)	46	43	27	23	14	57
Distributed computing/storage systems (e.g., LOCKSS or iRods)	4	4	6	16	19	21
Cloud storage (e.g., DuraCloud, Amazon S3, Google Storage, Mozy, or Box.net)	5	2	6	4	4	12
Other solution	7	5	10	6	8	15
Number of Responses	61	57	60	58	50	63

If you selected "Other Solution" above, please briefly describe the solution below.

Other solution for ingest N=7

Bagit transfer protocol.

Cloud storage is currently used on a limited basis for ingest; we plan to investigate its use for the other categories listed in this survey.

Consortium provides web-based ingestion, processing, and access for thesis and dissertations.

Currently all are being reviewed.

Hosted Open Repository.

Library IT runs a collection development instance of DSpace on its own server.

OnBase.

Other solution for processing N=4

Consortium provides web-based ingestion, processing, and access for thesis and dissertations.

Currently all are being reviewed.

Library IT runs a collection development instance of DSpace on its own server.

OnBase.

Other solution for access N=9

Amazon Cloud, hosted Open Repository.

Consortium provides web-based ingestion, processing, and access for thesis and dissertations.

Currently, all are being reviewed.

Local implementation of a Fedora repository.

Shared servers with IT on campus.

Local DSpace instance; California Digital Library's Web Archiving Service; system-wide open access repository.

We are still working this out.

We use OhioLINK for some digital content, not necessarily born-digital content.

YouSendIt & email have both been used to provide access to materials.

Other solution for back up N=5

Amazon Cloud, hosted Open Repository.

California Digital Library's Merritt Repository (content repository, geographically separate).

Currently, all are being reviewed.

Redundant storage managed by campus and library IT.

Virtual and physical tape storage

Other solution for long-term, dark storage N=7

California Digital Library's Merritt Repository (content repository, geographically separate).

Chronopolis.

Currently all are being reviewed.

Isilon.

Redundant storage managed by campus and library IT.

Virtual and physical tape storage.

We do not have "dark storage" per se. Instead we use Fedora as an asset management system where "master files" (e.g., tiffs) are copied to our replicated storage systems for long-term preservation, with appropriate preservation metadata and restricted access.

16. Please briefly describe how your library estimates future digital storage needs and costs. N=44

Analyze past usage and extrapolate future as well as monitoring use on a monthly basis.

Archives use virtual servers for digital storage, and adds storage as needed. Pays monthly fee to central IT. Estimate 5TB year. Libraries use a combination of local storage area network, remote storage area network, and offline tape backup. Estimates are based on current collection growth and future predictions for growth. Currently operate with 30TB headroom for approximately 30TB of data.

Assessing past growth rates, adjusting for known projects forthcoming in the next year. Estimates are also adjusted to incorporate the storage needs related to grant-funded projects. At this point, costs are estimated based on current costs for disks, storage devices, and backups. We are, however, looking at ways of moving to endowment-based models for some of our storage costs.

Based on current usage and growth over time. We currently have over 15TB online and 100TB in dark archive storage and have reports for growth over time.

Based on growth rates for past digital collections projects.

Based on projecting growth from current collections and rate of estimated future reformatting and ingest.

Can't answer, this is done by library IT.

Collection staff are polled regularly and asked for estimates of incoming born-digital materials.

Curatorial and project management staff submit estimates on a quarterly basis. The IT department then analyzes the needs and costs for budgeting and acquisition purposes.

Curators consult with our digital preservation officer and estimate possible future digital storage needs based on past needs.

Currently done by IT library staff. Anticipating using L.I.F.E. model for anticipating curation/lifecycle costs. Processing storage vs. long-term archival storage.

Digital Library Services staff provide yearly estimates on the growth of digital assets in the system. Estimated growth is determined through an evaluation of existing programmatic support as well as identifying particular projects that may bring in additional assets. A longitudinal analysis is also done to see how we are trending over time in terms of our digital storage growth. This information is presented to Central Operations staff for use in budgeting and storage acquisition decisions.

Estimate based on storage growth in previous years.

Extrapolating from current use and engaging with vendors/partners (i.e., CDL).

Future digital storage needs will be scaled to the development of campus department operations. The trick is to develop a system that is flexible, sustainable, and migratable.

Future storage needs and costs are managed through the ERA Program Management Office, who must balance the storage needs of all the instances of ERA against the most cost-effective storage approaches.

Have not yet.

Libraries IT solicits estimates from special collections/archives for the next year's usage and needs. Estimates are based on past usage and growth and anticipated projects. Anticipated projects may be either digitization projects or born-digital content we expect to receive.

Moved away from DVDs and external drivers (except occasionally for ingest) and work directly on server (IT supported Network File System) for all steps in the process. Storage needs: so far, have depended on the recommendations from DLI programmers and systems staff.

Needs are estimated based on known incoming materials in the short term, ideally several weeks/months in advance. We will gradually add TB of storage space as needed.

Not applicable at this time. Pending.

Our storage projections account for our born-digital and locally digitized materials and are based on the fact that we will have a number of file types which, both in their native and any normalized formats, are quite large: for example, uncompressed tiff files and video files, and large datasets. Costs are determined by library and central IT.

Our units look at recent activity requiring digital storage, and at future projects and goals to estimate our upcoming storage needs.

Past years' growth and projected new acquisitions of born-digital collections.

Planned digitization activities or acquisitions of born-digital material are planned for the fiscal year. Storage amounts required to accommodate that digital content are devised based on average file size for a particular type of record. Costs for this storage are estimated based on current market value, usually at the TB level.

Planning for digital storage needs and costs is the responsibility of the library systems department and the Associate University Librarian, Digital & Discovery Services, with consultations with department heads on their storage needs for ongoing activities and special projects.

Project by project, case by case.

Read research and follow trends. (Our recent move to Cloud storage for example. Once method is tested by other institutions and proven to be trust-worthy.)

Still developing.

The Digital Data Curator and Director of IIS monitor storage utilization and recommend purchases for grants and for annual purchase based on the types of materials currently stored and anticipated storage needs for project in planning or currently under way.

The library is working on a plan to estimate future storage needs now. Currently it is just allocated on an as-needed basis.

The recent inventory is a first step. In addition to the inventory curators have been asked to estimate growth rates. Library IT is investigating storage options including the cloud and cost models for storage.

This is a process receiving on-going development. Currently, space needs are estimated given past collecting volume + a 20% inflator and any known collections we anticipate receiving. Costs are estimated by the library IT staff based on the cost for the storage they lease from the university's Office of Information Technology.

Track historical usage and growth contrasting the resultant data with projections/requests previously provided from librarians. This provides a delta of growth not contained in a long-term plan. Track usage from new initiatives. Categorize the data by type; identifying growth areas. An example of historical usage follows: in November of 1999, the fileserver had 5 GB of disk storage available for all library employees. In 2000, there was approximately 30 GB of storage made available. In April of last year we were backing up around 10 TB of data. Currently we back up 18 TB of data. This data is "information" versus server images, etc. An example of a new initiative: The Dean has expressed a

strong desire to have all data replicated and online off-site which would bring the immediate potential usage to 36 TB. This amount has to be doubled due to redundancy on our ISCSI SAN (72TB) and multiplied by 1.17 to factor in RAID (84 TB). Plus, keep a 24TB node on site for redundancy (108TB). Since technology changes, we always start with the original amount of "real" data. Rule of thumb, consumption generally increases by a factor of 2 to 4 within a 12–18 month period. However, create a new department, get a grant, etc. and projections and planning is not quite worthless, but….. Prepare short term solutions for immediate growth needs (generally encountered through some event horizon effect). If a list of digital collections and their respective size estimates for the next 5 years were to be provided, more precise projections can be made- if no deviations to the plan are allowed. Baring that, any additional new (unexpected) collections should include monies for the growth in disk storage and allow for the delivery of the necessary hardware.

University IT storage fees plus staff time.

We are in the process of assessing storage needs for digital archives and University Archives over the next three to five years. We are basing numbers on collection growth expectations and assumptions about the types of media we will likely acquire. Costing models are established by University IT.

We do not currently have a metric for this process, but will be working to develop one.

We don't currently do this. Some staff understand that this is a problem, but few at the executive decision-making levels.

We estimate storage needs and costs based on past growth and known new projects and commitments. We also add estimates for possible and unpredictable needs insofar as possible. We do careful hardware and market analysis to determine best vendors, configurations, and prices.

We have a pipeline of projects, estimate space per project and negotiate with Central IT for space. The library's DSpace server acts as a buffer until production, enterprise space becomes available.

We have built an Excel spreadsheet that lists expected and 'prospected' collections and collaborative projects and their estimated storage amounts; and used formulas based on cost of Isilon, support, and staffing, for example: 10 hours/video = 1TB = $3000/5 years of storage on Isilon.

We plan on using DuraCloud and Peachnet (cloud storage) for future external storage and replication. Estimates are based on current storage needs with estimated growth rate of 4TB per year.

We project ingestion of electronic records residing on legacy media in the future, but have no way of accurately estimating born electronic records residing in systems like Banner.

We're still working on the best way to estimate future needs. Currently, frequent communication about upcoming projects helps estimate these needs.

STORAGE CHALLENGES

17. Please briefly describe up to three challenge(s) your library has faced in storing born-digital materials (e.g., amount of storage, security/ability to store sensitive data, ease of access to digital materials in long-term storage, cost, technical skills in setting up and managing storage, etc.) and how the library has addressed that challenge. N=57

Storage Challenges Word Cloud

Ability to store various kinds of data, ease of access to materials in long-term storage. We are working with an external consultant to establish digital preservation services, including long-term storage.

Accessibility of electronic records: Because of our need to search and access a heterogeneous body of electronic records at the individual asset level we needed a storage solution that supported indexing and searching for an array of formats. EOP ERA uses the Hitachi Content Platform storage architecture, which indexes hundreds of file formats, including the email and various office automation products that comprise the majority of the records in our holdings. Our deployment of the Hitachi platform includes the FAST search engine, which allowed us to develop search interfaces that correspond to the predominant types of records in our collections. Balancing ease of access with security. The amount of storage necessary to ingest and reserve federal agency records and the cost for the massive amount of storage that will be needed over the long haul.

Acquiring adequate volume of storage at a reasonable cost. Our A/V collections consume a large amount of space and all our materials require security and redundancy. Tiered storage solutions would provide more cost-effective solutions, however they are not available to us. We have materials in several locations and we do not have adequate means of tracking what materials are stored where. The coordination to increase network security around sensitive dark storage has been difficult. Miscommunication across four different groups for this new service has been a very slow process.

Administrative reorganization: Libraries is reliant on the university's IT department (IST) to purchase and set up digital storage. IST is undergoing a major reorganization making it very difficult for staff to be assigned to carry out the Libraries' requests for storage. A new CIO has been hired to finalize the reorganization process. Libraries has investigated cloud storage as an alternative to IST. Size of digital files: the Archives has acquired several TBs of born-digital records and has created several more through its ongoing digitization efforts. This fact, in combination with IST's difficulty in enabling additional storage, results in the Archives quickly exhausting existing storage options.

Centralization: our situation will likely be exacerbated by the fact that the Libraries' IT staff will soon be subsumed by the university's IT staff in its efforts to centralize functions. We can expect greater delays in acquiring and managing digital storage when we lose our dedicated IT staff who were previously responsible for these tasks.

All is in flux and subject to change, but we are developing a DAMS and staffing to develop that system is slender. We have one programmer on the job, making progress. Need to address security/privacy issues more fully than we have and develop comprehensive (rather than case by case) strategies.

Amount of storage: continually asking for more. Explaining how this is different than a preservation repository, which the materials will go into but until they are processed. Ensuring stability of the files.

Amount of storage: current university infrastructure does not have capacity for a large amount of born-digital material. Future upgrades should take in to account an exponential increase in expected storage need. Secure access: For those items we choose not to or cannot store on campus, choosing a cloud-based solution is difficult because of PATRIOT Act issues. This is an ongoing issue. Staff expertise: IT staff are not necessarily versed in maintaining archival quality records. This is primarily a staff training issue, not a technological one.

Amount of storage: We nearly ran out of space this year due to the way the servers were configured and allocated. The problem was that a server had been called into service to temporarily house a system from a failing server. This issue was temporary, as the system is being migrated to a new server, but it is indicative of space budgeting problems. We have not always been accurate in our predictions of space needs. We have recently moved to a VMware solution that should help by providing greater flexibility. Cost: This has historically been a problem. Initially, collecting areas that took in or created digital content were expected to pay for their own servers, but in recent years this has become an accepted part of the Libraries IT department's responsibilities. Our costs have also gone down as a result of a move to VMware. Technical skills: most recently this has been in the area of awareness of the need for (and skill in integrating) things like integrity checking and monitoring systems in general. This will be the next step in special collections/archives' collaboration with Libraries IT staff.

Amount of storage has previously been a challenge, but has become less of a problem with the fall in storage costs in recent years. Future storage needs for large-scale ingest of born-digital special collections materials will probably be integrated into university-wide planning for digital repositories, a digital asset management system, and networked storage & continuity planning. Technical skills of special collections staff in managing born-digital materials has been a challenge, which was initially solved by contract staff with the required skills, and now by hiring a Digital Special Collections Librarian as permanent staff with the required skills. Providing access to born-digital special collections is an ongoing problem, with no unified solution. ETDs are available through a DSpace instance; the university web-archive is hosted externally, with Archive-It. Copies of other born-digital materials in special collections or university archives fonds are usually provided to researchers on a cost-recovery basis, using optical disks. Future development of a library digital repository will greatly facilitate access to the latter materials.

Amount of storage needed and "non archive" approach of central university IT unit. Still under discussion.

Amount of storage required and costs of storage. Staff resources and funding for managing born-digital records. Time resources for those with technical skills for storage management.

Amount of storage required. Cost.

Amount of storage: working with library IT to provide server space.

Amount of storage; starting to manage temporary alternative file management systems. Ease of access; challenges of ongoing equipment management. Technical skills commitment from institution.

Amount of storage: We have purchased additional disk space. File type/migration: We are in the process of creating a digital preservation policy to limit file types we will manage/migrate. Security: We have developed project teams with limited levels of access, depending on need.

Cost. Distributed storage sites. Long-term storage and access.

Dark archives storage is less of a preservation environment than the platform for access copies. Storage issues are further exacerbated by a lack of central IT understanding of digital preservation requirements. The previously mentioned inventory is the start in a process to get a better handle on storage capacity and digital preservation tool needs. Further, the Libraries have identified the need to develop a digital preservation policy/plan. Staff time and skills to actively ingest, process, and manage born-digital objects. The creation of the e-records/digital resources archivist position is one step in the process. Additionally, curators and curatorial staff are seeking appropriate training opportunities.

Data loss: Moved toward more stable hardware and regularized review. Technical skills: Hiring consultants as well as using combination of library and parent IT.

Disconnect between archival masters, metadata, and access derivatives. Previously these have been in separate systems or in a simple file system. We are implementing a Fedora-based repository service to centralize storage and management of ALL digital materials. Managing rights and access to restricted content. We are working with university IT to implement Shibboleth identity management as one approach to solve this problem. Determining the long-term cost of storing digital content in perpetuity. We are examining pay-once-store-forever vs. subscription management models.

Funding and skills to manage a true digital archive. Amount of storage required.

Having enough storage so that we don't run out during a project. Getting an appropriate system in place for off-site, secure back up. Cost of storage.

High cost of preservation storage infrastructure. This has been addressed for the present by reallocating funds from other parts of the Libraries budget to purchase storage. When feasible we add a one-time storage fee to grant-supported projects. Bandwidth costs. Because of bandwidth costs, we have selected remote storage options that are available via subsidized carriers like NYSERNET or Internet2. These storage options are not necessarily the most cost-effective, however. Changing storage technologies, manufacturer, and vendor churn. We have approached the problem of vendor churn and changing technologies by assuming a rolling five-year model for hardware replacement, assuming that we may have to keep changing vendors and equipment.

I have been delaying moving digital video from media to server space because of the massive file sizes. We also have issues ripping video from certain access formats. I am still requesting MPEG-2 for film/video preservation since it is compressed; don't have time to evaluate sustainability of other formats that might enable increased quality.

In terms of personal materials we have had issues related to ease of access and managing storage. But all of these are potential issues since our current system for personal materials is sketchy.

Insufficient staffing. We continue to explore options within the context of library-wide staffing issues. Long-term storage and access. We continue to work with library IT and university IT. The actual amount of storage needed. We continue to work with library IT and university IT.

Justifying the need for and the resources required for storing multiple copies of large original/master files, in multiple locations, and preserved on an ongoing basis. Affordable, geographically distributed storage. We are evaluating options to distribute storage of our digital materials, mitigating the risks associated with a single location for storage. Costs of large-scale storage. We are reexamining our business models for the storage of digital collections and investigating partnerships that will make it easier for us to manage storage.

Lack dark archive storage structure and toolset. Technical skills to manage stored content appropriately (are currently working with Fedora and associated management tools). Space: cost of long-term storage, especially for content from researchers.

Lack of preservation-quality digital repository. The library has designed and implemented a Fedora-based repository system to serve as a dark archive. Inexperience managing digital files on a server (setting up file structure, etc.) We will collaborate with other library groups to establish consistent best practices. Safe handling of donated servers. We are investigating best practices for handling these materials.

Larger institutional inertia on electronic records management has not yet been addressed. Lack of "one size fits all" solution lowers curatorial enthusiasm for managing born-digital records. Technical solutions/skills/infrastructure also hasn't been addressed yet.

Learning what level and type of metadata must be preserved to accompany the growing amount of born-digital assets, including: adequate Dublin Core records for collections in DSpace; embedded digital metadata; workflow and seeking ways to efficiently transfer existing metadata; how to address concerns about linking to resources and the possible transient nature of links. Selecting versions of materials to be preserved. For example, should files be saved as originally named as received and in the form re-named for local use? Should all formats of images be saved – tiffs, and any derivatives; or wav files and derivatives, or only uncompressed formats? In a way the question is whether we are storing for preservation or to provide an inventory of formats for delivery and service. Archives are guaranteed preservation only if stored on enterprise data storage. Redundant, highly available enterprise disk is still costly.

Long-term storage, backup & mirroring, geographic distribution of mirrored sites: Libraries has worked to include infrastructure expenditure into operating costs. Availability, access restrictions, copyright: library has met these challenges on an ad-hoc basis.

Network Bottlenecks: Moving large amounts of data across our network has been challenging due to bottlenecks which result in failed process and excessive transfer times. This is an issue that we are currently assessing. Storage Capacity: Our present storage capacity has not kept up with the rate of acquiring and generating born-digital materials. Our institution is presently developing a preservation repository that will have increased storage for items that we are interested in keeping in perpetuity. Procuring storage is also difficult. A number of library stakeholders with an interest in digital content will be working on a committee with a member of library IT to address workflow issues that may improve storage efficiency. Security: We are encountering challenges with providing access to materials that are subject to copyright. Although we are reformatting items as deemed acceptable under Section 108 we still have to protect these items from illegal duplication. Thus far, it has been difficult to provide access to these items.

No integrated digital acquisitions plan; planning tends to be on a project basis rather than an overall program. Capacity challenges within our physical technical infrastructure. Ability to manage access based on a wide and changing variety of licensing and access restrictions.

Not enough available server space for storage. A larger server has been purchased. Lack of staff expertise regarding born-digital material storage. This has not yet been addressed.

Our biggest challenge right now is in storage capacity, given the fact that we resort to an array of external hard drives. We are developing a pilot Digital Asset Management System that would expand our capacity. Second would be the need for developing and implementing metadata to effectively address records retention schedules. Long-term storage is the final challenge. Our plan is to develop an effective DAM system with an archival system to ensure preservation and access.

Our biggest challenge with storage continues to be establishing policies and infrastructure that will allow us to integrate born-digital collections into our larger repository infrastructure. Security, complexity of ingested items, and size of objects have been impediments. We are taking a phased approach to ingest and working with our systems staff to address these challenges.

Our storage is currently for archival resources only. Research faculty would really like storage where they can build research data, by continuing to add and revise data until it is ready for permanent ingest. We are currently looking at strategies to segment our space and provide work area utilities. We currently have 47 TB of data storage, which would not be adequate for very large data projects. We do not have a successful working model for assessing the cost of data storage, which we believe needs to be a one-time cost but must provide at least partial cost recovery for managing data over the long term. Most models we have seen are based on the cost of storage, not the cost of staffing for storage, which should include the cost of preparing and describing data.

Quantity of storage available, including appropriate backups. Cost of increasing amounts of storage. Setting up, monitoring, and managing increasing amounts of storage.

Security of sensitive material. We are investigating the ability of Rosetta to segregate materials and allow access by user password. Access to born-digital materials by patrons. We are trying to determine if we need to have a public access system and a dark archive or if one system can do both. This is contingent on solving Challenge 1. Funding storage costs. We are working with the university administration to see if they will fund some storage costs and we are investigating a model where we would grant campus departments a certain amount of space and if they need more, they would pay for it.

Security/ ability to store and manage sensitive data (work in progress). Policies to address storage requirements (work in progress).

Sensitive Data: No Research Data has yet been made public. Access is restricted to the research teams. Our Graduate School has also declined open access to ETDs. Repository ETDs are restricted to staff, but the public can access some through ProQuest.

Server space and the management of that space is a challenge. We have set regular meetings between staff in the unit using the largest amount of storage space and the library IT staff to be sure all are informed on upcoming storage needs.

Storage space and estimating storage space. Coordinating storage. Fixidity.

Storage space that is not an external hard drive in someone's office. Still trying to figure out how to handle this. Cost: which system is the most economical but also does what we need it to do. Still trying to figure this out. Access to born-digital materials. Not sure yet.

Storage space. Library IT purchased new servers and is collaborating with campus IT for additional storage space. Restricting permissions to specific viewers is challenging on an administrative level. Digital Initiatives and Data Curation Librarians work with data providers to determine who should have access to data.

Technical requirements in setting up a sustainable digital preservation environment. The Libraries is continuing to define all the aspects that make up a fully functioning preservation environment, looking at the needed policies and procedures, application support and technical infrastructure. Any final policy or plan must fit within and be driven by existing Libraries collecting policies. Issues related to long-term sustainability of assets in a multitude of formats, some standardized and others of a more non-traditional or uniquely proprietary nature. The Libraries is looking to push out support for those submitting files for inclusion in its systems with recommendations for file format types that are more easily sustained or that have proven better for support.

The main challenge is that our existing data that we use to inform storage needs are based on the creation of image collections. Born-digital materials have the potential to be exponentially larger in terms of storage requirements. Distributed storage environments. We have not yet identified a sustainable way to hook our repository services up to other campus storage environments for the purpose of linking and ingestion. Estimating the costs to maintain storage for the long term, including curation and migration costs.

The most significant challenge for storage is the same as one of the challenges for ingest: the lack of an infrastructure of repositories and tools to store and maintain these kinds of materials. While many parts of the process can be handled by established tools, other parts can't or systems don't work together. So, as an example, while we have preservation storage for digital masters of digitized images that can also be used for storage of born-digital materials, this is dark storage and it doesn't meet the need for access and discovery of materials. Similarly, the repository for access and discovery has been designed so far for individual images, e-books, and A/V, not for heterogeneous groups of materials in a manuscript collection that may be described only at an aggregate level in a finding aid. The only way to address this challenge is to develop the infrastructure further and adopt and adapt emerging tools for parts of this process. Related to this is coordination of resources to address these needs. While we, like everyone else, can always use more staff and more funding, just utilizing the staff that we have to address these needs while continuing to address existing needs is a concern. In addition, the infrastructure and workflows created to address these materials cannot exist in a vacuum, they must be compatible with or must be an extension of the infrastructure that manages data about the rest of the library's collections. This means that progress on developing infrastructure in support of born-digital materials must include input, buy-in, and resources from many parts of the library organization: special collections, library IT, administration, technical services, etc. Increasing meaningful communication between groups and jointly planning development is the best way to address these challenges. Another challenge for storage is determining what to store and what metadata to store about it. While a default option is to store a bit-level copy for long-term preservation, some work has been done to determine what other levels of preservation can be supported and what data would need to be stored in order to enable this level of preservation. A significant challenge relates to the retention of private or sensitive information. Given the nature of the archival workflow, we really do not have time to completely process materials before we put them in archival storage. This means that private or sensitive information may be inadvertently stored for some time. We do have some tools we can use during accessioning to automatically search for significant patterns such as SSNs and social security numbers within textual data, but we do not have the time to do any more in-depth searching. While this mirrors the situation with paper records (although, we can actually remove more potentially private information during accessioning with born-digital textual materials than with paper), the risk is much greater for loss of security of this information in the digital environment. An additional issue that we are still considering is how to (or, indeed, whether to) securely dispose of media carriers (disks) that continue to store sensitive data even after a copy has been retrieved from them. The issues are that, in some cases the media carrier itself may retain artifactual value (hand-written annotations, modifications, metadata contained on labels, etc.), if the copy made was corrupt or lost the media can serve as a back-up, and that completely wiping hardware is difficult to do. The recommended options for destruction and deletion of the data are potentially costly and time-consuming (disk shredding, magnetic wiping). To address these challenges we have undertaken a number of activities and are still discussing other solutions. One major step was consulting the library's legal counsel for advice on adhering to university, state, and federal regulations in the handling and storage of this data. Other workflow issues, such as the screening of content for sensitive information at the accessioning stage using automated methods, have also been added to the workflow.

UCISpace Fixity: until recently the material ingested into UCISpace (local DSpace instance) was not being continually checked for fixity/authenticity. We now run a checksum checker on all UCISpace content nightly and are in the final stages of implementing a system that will back up DSpace generated AIPs of all UCISpace material into CDL's Merritt repository. This Merritt collection will serve as a geographically separate dark archive that we can also access to replace lost or corrupted items/collections if and when the checksum checker discovers them. Canto Cumulus: a robust digital

asset management system that our Special Collections and Archives department uses for managing media collections, mostly digitized and born-digital images. However, it has a steep learning curve and is not very user friendly, and we have had difficulty obtaining vendor support for the product. When Special Collections acquired the system, they had more staff available with responsibilities for using the system. However, due to staff attrition, remaining staff cannot devote the time required to learn and utilize the system effectively. We are open to using an alternative digital asset management system that is supported by the entire University of California system.

Unsure of what long-term costs are. Unsure of where to put the materials. Gap between best practices for digital preservation and current storage method.

Very little systematic thinking. There is no single person or unit responsible. A number of different people and/or units have been responsible for mass storage that may be utilized for long-term storage of born-digital collections. This is not a strategic priority for the institution. Mixed content. Mass storage includes important born-digital collections, surrogates of digitization projects that may or may not have long-term preservation value, and other mixed content that has not been appraised in any way. Cost carried entirely by collectors. Unlike paper storage, which is a shared library expense, digital storage expenses are allocated by the collectors. There is no current budget model that allows for the sharing of storage sufficient for born-digital collections. This is particularly a problem for special collections.

Volume of storage, we've added capacity to the system. Access to digital content is provided by DSpace. Much digital content in MASC is staff only access.

We are establishing an e-records workstation in a locked office with a secure connection to the dark archives server. Original media retained with PII will be stored in the vault used for rare books.

We benefit greatly from challenges and benefits of scale. We have local, campus-based cloud storage through centralized IT (CNS) that gives all of the benefits of cloud storage with no negatives. We are able to leverage capacity for maximized benefits.

We can't seem to get enough storage from the central IT units, and the storage we do get is doled out to us in relatively small chunks.

We have had to increase storage capabilities by working with our main IT department on a regular basis. (Particularly how our storage needs have grown substantially since first incorporating the institutional repository.)

We need to establish our preservation policy. Multiple formats and file/format stability.

TOOLS

18. What software/services/tools does your library currently use or plan to use for digital processing actions? Check all that apply. N=54

	Currently Use	Plan to Use	N
Open source tool (e.g., Jhove, Droid, SENF, ADAPT ACE)	31	13	44
Outsourced service (e.g., Archive-It)	12	19	31
Home-grown tool	18	11	29
Commercial tool (e.g., Aid4Mail, IdentityFinder, etc.)	21	6	27
Other software/service/tool/approach	8	5	13
Number of Responses	42	30	54

Please list the specific tool and/or briefly describe your approach (from bandaid/bootstrap approach to microservices software development) below.

Commercial tool(s) N=22

Adobe Bridge; Photoshop

Archivists Toolkit, CONTENTdm

CONTENTdm and Shared Shelf currently being used.

CONTENTdm (2 responses)

CONTENTdm for access

Currently use Adobe Pro for conversion of some documents to PDF and PDF/A, and can anticipate using other commercial products.

Forensic Tool Kit, Aid4Mail

FRED, FTK Imager

FTK; FTK Imager; ImgBurn; Aid4Mail; DVD Decrypter; Md5Checker; Catweasal ImageTool3; FC5025 Imager; ArchiveFacebook; JR Directory Printer; MediaJoin; Quick View Plus; SyncBack; Kryoflux software; PCMacLan

FTKImager (though free, proprietary), FileMerlin, Oxygen, Adobe Acrobat

Hitachi Content Platform, FAST search engine, alfresco (for our users' work environment)

Identity Finder, McAfee Anti-Virus, FTK Imager (free), EmailChemy, and QuickView Plus

IdentityFinder

Isilon enterprise storage

Mac-legacy versions of iMovie hacked into earlier versions of product (all built into the Mac OS).

OnBase by Hyland Software

Photoshop, Acrobat

Quick View Pro, FTK Imager

Symatec Backup Exec.

Use Identity Finder for PII.

We are in the very early stages of exploring the potential of SharePoint.

Open source tool(s) N=34

Archivematica (2 responses)

Archivematica (incorporates Jhove, and other open source tools)

Archivematica for preservation, ICA-AtoM for access, DSpace for access.

Archivematica, BitCurator, Hypatia, Fedora

Archivemedica: thinking about it.

Archon and Fedora

Currently, DSpace, JHOVE and GIT

Droid, Duke Data Accessioner

DROID; FITS; LOC Bagger GUI; Bagit library; Thunderbird; Handbrake; IrfanView; Shredder; WinHTTrack; YPOPs; Fiwalk; Sleuthkit; MediaInfo; Afflib; ClamAV; Fido; Archivematica; Basilisk II

DSpace currently being used; Fedora, Islandora being considered.

DSpace, ClamWinAV, JHOVE, DROID, SleuthKit

DSpace, Drupal, FEdora, Solr

DSpace, Drupal, Open Journal System

Duke Data Accessioner, Archivematica, Fedora Commons, Jhove, FITS, IRODS

Duke Data Accessioner; also looking at Archivematica.

Duraspace (Fedora Commons), ExIf tool, MediaInfo, Archivists' Toolkit, Oxygen

Evaluating Duke Data Accessioner; Archivematica; California Digital Library's Merritt; MetaArchive, etc.

Exploring archivematica or DPSP from NAA.

Fedora Commons digital repository + Hydra/Blacklight

Hydra/Fedora tools (DIL), MDID3

JHOVE, BagIT, Archivematica, AIMS, FITS, fiwalk, LOC-Bagger, Curator's Wookbench

Jhove, Droid, Archivematica, Virtualbox, HTTrack, Imgburn, CDCheck, Exiftool, + others (whatever helps/works)

Jhove, Droid, Rosetta

JHove, ImageMagick, Libraries from open source projects such as Islandora, Emory's Fedora libraries

Jhove; check sum generator; GIMP; IrfanView; Heritrix crawler; Wayback Machine web archive replay software.

Jhove; Droid

NARA File Analyzer, Duke Data Accessioner

Not yet decided

Our digital asset management system, Islandora, makes use of several open source tools. We will also be exploring how we might integrate it with Archivematica.

Sheepshaver for emulation

Starting to use Archivsts Tool kit, some use of DSspace.

ThinkUp, Yahoo2Mbox, Thunderbird, JackSummer, ADAPT ACE, aimage

Win SCP

Outsourced service N=22

Archive-It (7 responses)

California Digital Library services (Merritt repository, eScholarship, Web Archiving Service)

CDL Web Archiving Service

Considering California Digital Library Web Archiving Service for web preservation and CDL's Merritt platform.

Eventually Hathti Trust

Internet Archive.(2 responses)

Looking at Archive-It for harvesting websites.

Looking at ways to archive websites.

MetaArchive Cooperative: members, but have only used it for one collection so far and can't afford to put everything there.

Not yet decided.

Outsourced disk imaging.

Plan to contract with CDL Web Archiving Service this year.

Plan to use Archive-It's web archiving service.

Several under consideration for archiving websites.

Web Archiving Service

Home-grown tool(s) N=18

Archive-It

Bag-it; Content Transfer System (CTS); DigiBoard (nominations & permissions tool)

Check-in/ingest tool (Medical Center Archives only)

Curators' Workbench developed locally and incorporates some open source tools.

ETD Processing system, various BASH/XSLT scripts

Exploring UNC's Curator's Workbench.

Homegrown tools (more bootstrap than microservices) to interact with Fedora and use on Macs.

In the process of development.

Metadata wrapper for storage and access of digital items in Fedora.

Python-based microservices and APIs

Scripts to support accessioning and metadata extraction; Media log.

Still coming up with this maybe.

Tools for managing repository and storage network workflows.

Various Python & Perl scripts

We are customizing the open source DAMS, Islandora, to meet our requirements. See above.

We could conceivably explore home-grown tools/solutions with our university IT staff.

Workflow Management System (metadata and object handling for RUcore)

XSLT stylesheets; Schematron and RelaxNG Schemas; local scripts

Other software/service/tool/approach N=13

Archivist's Toolkit

Bagit

Considering Archivematica.

Currently exploring a variety of tools, including: checksum checkr; metadata extractor; DROID; Xena; FTK imager; Adobe Bridge.

Hand encoded MODS records

Heritrix (Web Crawling)

I do not understand the question, which indicates the problem. We work with library IT folks to whom this would mean something, hopefully.

Open source cont: DROID, Duke Data Accessioner, Apache Tika, etc.

Photoshop, Thumbs Plus, and a variety of others to carry out specific processing actions such as mass renaming; addition of caption information to each digital image lacking that information; stripping out metadata that is embedded in the header of photos and creating a text file; tools to make mass corrections of file names (for example removing all empty spaces in filenames).

SIARD

Special Collections staff personal collection of archaic software & hardware.

We acquire/develop tools as needed.

XTF, Omeka, Digital Commons, DLX, Archivists' Toolkit, Silverfast, Photoshop/CS suite, Adobe Acrobat Pro, ImageMagick, Tesseract, iMovie, Final Cut, OmniPage, ABBYY Finereader. Not sure what is being asked for this section.

ACCESS AND DISCOVERY

19. Which of the following delivery methods does your library use to provide access to born-digital materials? Check all that apply. N=64

For the purpose of this question, in-library access refers to a reading room or other monitored space; online access means access to materials remotely; i.e., not in a monitored space.

Online access to a digital repository system	42	66%
In-library access on dedicated computer workstation	31	48%
In-library access using portable media accessed through the users' personal computer	22	34%
Third-party access & delivery system	18	28%
Online access to a file space	15	23%
In-library access to records in an emulated environment	1	2%
Online access to records in an emulated environment	1	2%
We do not provide access at this time	13	20%
Other delivery method	10	16%

If you selected "Third-party access & delivery system" above, please specify it here. N=14

A small number of born-digital materials are included in our online CONTENTdm system, e-Archives.

Archive-It provided portal & YouSendIt

Archive-It.org, for the university web archive.

Campus-based Dropbox file sharing to send large scanned documents to distance researchers.

CONTENTdm (2 responses)

CONTENTdm for selected collections – derivatives only.

Digital repository content is syndicated to a number of online systems, including the library's VuFind catalog.

Dropbox, YouTube

LUNA Insight, CONTENTdm, ViewShare

No public interface to digital archives, so staff must provide requested digital material to researchers.

OhioLINK, but not strictly for born-digital. And once again the typical born-digital assets in OhioLINK are ETDs which are not under the purview of special collections.

Use of vendor sites for access/delivery of purchased content. Trusted partners that host parts of our digital collections.

We are in the early stages of exploring access via our Bepress/Digital Commons-based institutional repository.

If you selected "Other delivery method" above, please briefly describe it here. N=10

Ad-hoc digital libraries

Creating duplicates for patron use on storage media (CDs, DVDs).

Digital documents & images delivered to users as email attachments.

Email

Home-grown PHP app: customized file/directory browsing application

In addition to our plan to make records available online through OPA we can also deliver .zip files of assets and metadata to requesters.

We deliver all through online access whenever possible. In some cases, as with partners in different areas and countries with limited bandwidth and with materials of varying levels of extreme sensitivity, we support other modes of access as needed.

We generally do not provide access at this time, but in rare instances have provided access at a dedicated workstation in the reading room of the archives.

We pull DVD's and CD's.

We send files by the university's digital "drop box" and by email.

20. What repository system is used to manage and/or provide access to your library's born-digital materials? Check all that apply. N=63

	Manage	Provide Access	N
Open source repository software (e.g., Fedora, Archivematica, DSpace, or DAITSS)	39	33	41
None, the library uses secure file system storage	28	10	29
Commercial repository product (e.g., Rosetta)	10	12	15
Home-grown repository system	12	11	13
Other repository system	7	7	9
Number of Responses	60	53	63

If you selected "Other repository system" above, please briefly describe it here. N=14

Manage and Provide Access

CONTENTdm. Archivematica and ICA-AtoM being piloted.

Hathi Trust

Institutional records: OnBase. Personal materials: We are simply storing at the moment and not providing unmediated access due to lack of a repository for storing this material.

Shared Shelf

The original preservation copies of born-digital records are stored in a secure content management environment, the Electronic Records Archive. Those original copies are managed in archival storage and not accessed by the public. We make reference or public access copies of born-digital records and provide access to them either on hard media (for direct reference requests) or place the access versions of the files on the Online Public Access (OPA) web servers. Users can then search OPA and view and/or download the reference copies of born-digital records.

Manage

Fedora repository is under development.

Provide Access

Bepress Digital Commons

We use CONTENTdm for access. While this isn't always considered a repository system, this is the nearest fit on the survey. (We also referred to it as a repository system in question 14.)

Other Comments

CONTENTdm is used now to provide access to select born-digital materials; however, it is a short-term solution as we evaluate platforms such as Merritt that can provide both preservation and access.

Duraspace (Fedora)

Medical Center Archives does not currently have a digital repository system, but development is underway.

Open Source repository software with SobekCM

VuFind with Solr; Active Fedora stack with Blacklight, Solr

We are currently transitioning from a secure file system storage with no access to a Fedora-based repository system.

21. Are different types of repositories used for different types of born-digital materials? N=60

Yes	38	63%
No	22	37%

If yes, please briefly describe which type of repository is used for which type of material. N=36

A locally created web-based access system is used for the University Curriculum Archive, which is a mix of digitized and born-digital content.

Archival electronic records both via file system server and DSpace.

Archive-It serves as a commercial repository product for web-archived material. DSpace, and open-source repository software, is used to provide access to some open materials that can be described at the item level. The rest is managed via file storage.

Bepress (commercial) used for institutional repository; open source plans for Special Collections.

CONTENTdm is used for searching and access. DSpace is used for dark archives (the Institutional Repository).

Currently, two instances of DSpace are used to deliver some born-digital content.

DSpace for thesis and dissertations and some university electronic content; a secure limited access sever space for digital content, such as mpeg movie files, created in MASC digitization projects.

Described above: Archive-It for web archive; DSpace for ETDs; file system storage for other born-digital special collections materials.

DSpace for GIS data; CONTENTdm for documents, audio, visual materials.

DSpace for ETDs. Shared Shelf for photographs and artworks.

DSpace handles our institutional and subject repositories and is primarily a text and data focused site. UMedia Archive (Drupal, Fedora) is used to manage and present our rich media and image files.

DSpace is currently used for the output of the university's research community, e.g., theses and dissertations, datasets. Islandora, our digital asset management system, is used for all other forms of digital content and will be utilized for born-digital records too. Archivematica will also be investigated with regards to processing and managing born-digital content with ICA-AtoM providing access.

DSpace is used as the platform for our institutional repository; we will be using a Fedora-based system for our long-term digital preservation repository.

DSpace is used for our institutional repository; CONTENTdm is used for material digitized in the library; ICA-AtoM is used for collected born-digital material.

DSpace is used for traditional digital objects that DSpace usually manages. Everything else is not in DSpace.

DSpace: theses, DTDs. Homegrown: digital photographs, film, etc.

Fedora is used as basis for dark archive for all materials. DSpace is used as repository and for access to mostly textual materials. Other systems are used for access to images, video, etc. Materials will be stored in Fedora-based repository.

Fedora system used for digital collections and images; VuFind - selective content including some e-text;

For institutional records we are using OnBase an ECM system. We are still figuring out how to handle personal materials. OnBase may be our solution, though likely it will not be.

Images in LUNA Insight, documents in DSpace.

Institutional repository.

Manuscript vs. University Records.

Most files are kept on disk space managed by the Carolina Digital Repository, but extremely large or numerous files are kept on a tape-based storage system.

Not within Special Collections, but other units in the library are using other systems (Bepress, Luna).

Secure file system storage and Fedora are for dark archive material and deep storage. Fedora use is still in test stages, so it may also be used for access copies of digital content at a future time.

The institutional repository (DSpace) manages simple files. Secure file systems storage is used for complex datasets with access through DSpace. CONTENTdm manages curated special collections.

There are three repositories currently in use, although none of them are recommended for preservation, only access. The Libraries uses CONTENTdm for access to born-digital archives and special collections (e-Archives), the NanoHub for access to born-digital faculty research data sets (PURR-Purdue University Research Repository), and Digital Commons from Bepress (e-Pubs) as an institutional repository for access to faculty research articles, pre-prints, electronic theses and dissertations, etc.

Theses and dissertations repository provided by library consortia uses DSpace. Everything else managed on original media and/or file system storage.

We can't put everything in our Access digital repository. Collections such as copyright protected sound recordings, audiovisual material, sound files are only available on a case-by-case basis in-house or through limited time hosting platform (Omeka exhibit).

We have Scholar Commons for access to faculty documents that are born-digital and we have CONTENTdm that might be used for born-digital library collections, but has not really yet. There are a few films and oral histories in CONTENTdm, but that database is not for preservation, just access.

We manage preservation and access of books digitized by Google and Microsoft in Hathi Trust.

We use a Digital Asset Management system to catalog and manage multi-media material (mostly photographs and some audio and video). This system is for back-end use only. We export from it to other delivery systems as appropriate. We use a DSpace repository for our text-based, born-digital archival materials as well as for some content exported from our DAMS. The California Digital Library's Web Archiving Service is used for managing web-based content.

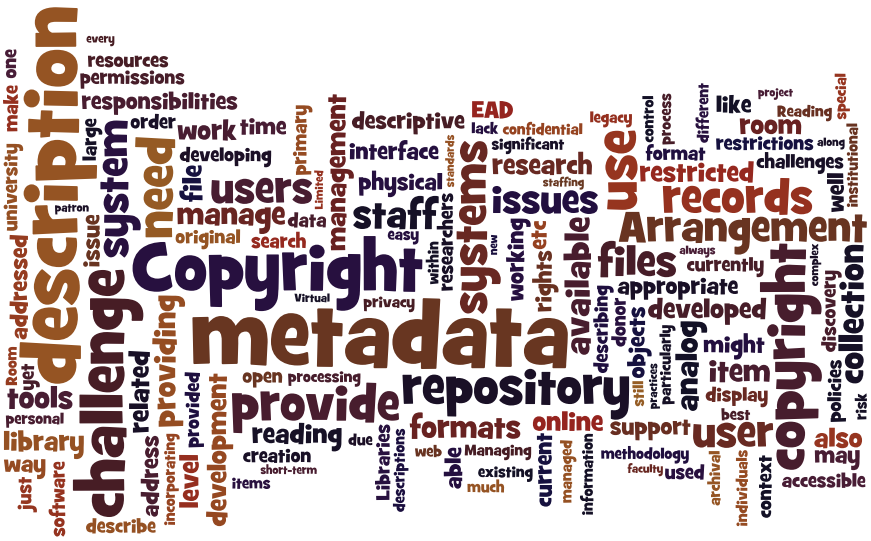
We use an institutional repository to manage scholarly content, CONTENTdm for managing our digital collections and OJS for managing journal content.

We're moving from scattered RAIDs, servers, etc. to the Isilon.

While we don't currently use different repository systems, Special Collections/University Archives plans to purchase an electronic records management system in the near future, which will probably be a commercial system independent of the library's Fedora repository. However, records of scholarly value that can be made openly available will be shared with RUcore.

ACCESS AND DISCOVERY CHALLENGES

22. Please briefly describe up to three challenges your library has faced in providing access to born-digital materials (e.g., arrangement and description, copyright, confidential content, etc.) and how the library has addressed that challenge. N=53



Access and Discovery Challenges Word Cloud

Ability to manage access based on a wide and changing variety of licensing and access restrictions. Technical infrastructure to manage access to restricted content beyond our IP range has not yet been developed. The policies and tools for authenticating off-site users have not been developed.

Adopting/developing descriptive standards and standard workflows for describing born-digital materials. Integrating descriptions of born-digital materials with analog materials. Handling confidential materials with appropriate access controls.

All manual process. We do not have a digital repository system that enables us to provide access to born-digital collections over the Internet. Most access must be provided on-site on non-networked computers. This is an inefficient, cumbersome, and slow process. Skilled staff are responsible for setting up computers upon patron request. Limited description. At this point, only some staff are comfortable describing born-digital collections. This leads to limited description of those collections that may be limiting use. Also, lack of appropriate tools to undertake description, particularly in the past. There is no infrastructure to manage user permissions (if there were an online access system).

Appropriate platforms to manage and deliver content (work in progress).

Arrangement and description of born-digital materials is an ongoing challenge, since current practices were developed primarily to deal with non-digital materials. Our new Digital Special Collections Librarian will be tasked with developing best practices for arrangement and amending current description practices to suit born-digital materials. Provincial privacy legislation creates challenges in providing access to born-digital university archives. This challenge may be addressed over time through implementation of a university recordkeeping metadata standard, which will facilitate

management of university archives according to the security classification of these records. The previously described issue of using multiple systems to manage and provide access to born-digital materials is another challenge, which will be addressed through future development of a library digital repository to rationalize storage and access, and also federated search tools to facilitate searching across multiple systems, when necessary.

Arrangement and description, processing.

Arrangement and description are the primary challenges to access and discovery. Born-digital materials arrive on legacy media with scant metadata to inform development of effective finding aids.

Arrangement and description of legacy material is a challenge because the media was managed as a physical item and arranged into one series, when the content may intellectually belong to a number of different series. We need to modify our gift agreement to clarify what kind of online access we can provide to born-digital material acquired from external parties (web access, library-only access, etc.) and also address the technological challenge of restricting access.

Arrangement and description; technical skills commitment from institution. Copyright and privacy; lack of policies and procedures. No good sustainable delivery mechanism.

Arrangement and description: collecting particular metadata up front; knowing what to collect (what subject experts or users might want plus what programmers will need—and how to crosswalk those elements); much manipulation of web display elements. File management-naming standards, organization, quality control, migrating files from DLI to Tech Services to Systems Programmers to Archive. Discovery and searchability of our digital collections, including Trace repository—OAI, OCR, finding aids, etc., as well as copyright issues.

Arrangement and description. The library provides simple searchable metadata records through the institutional repository. However, not all metadata is represented in these records.

Confidential content. We have records that are restricted for up to 10 years by the donor and have closed the entire collection until we are able to provide access only to the open content in a manner in which it cannot be altered by users. Copyright. We may not wish to make the full copy of an item available, or to make it available at a useable resolution. Remote access to large files. We've used Dropbox in some instances.

Copyright. Arrangement and description—currently focused on developing program.

Copyright: collection donor does not have copyright over content. Arrangement and description is not in line with the analog part of the collection, it is done separately and sometimes well after we have provided access to paper based materials. Time: we often focus on digitizing collections and providing access to those before we can work with the born-digital content.

Copyright: attempt to reach agreements with providers/publishers. Creating relevant descriptive metadata: metadata librarian supervises student workers. Development of access interface: Libraries are piloting Islandora, Archives are in the process of developing access interface.

Copyright and confidential content. We are investigating the applicability of the "one item one user" model that would limit access to copyrighted material to one authenticated user at a time. Similar to checking out a book or document in the reading room. Levels of granularity. Users expect item level access (or beyond) how do we describe this content in a meaningful way? We are exploring automated metadata creation tools such as document analysis.

Copyright and licensing. Consistency in user entered metadata.

Copyright: we use systems that allow very granular control of permissions and access. Privacy: we have policies governing access to confidential records, but procedures specific to born-digital materials are still being developed. We hope to use systems that allow very granular control of permissions and access.

Copyright: copyright statement in tagged metadata. Software upgrade/migration: purchasing new software and hardware. Limiting access to private collections: through software, some collections are only accessible in-house or through restricted IP.

Copyright: might actually be less of an issue than with some of our legacy analog collections, but people send digital content without deeds of gift just as they do analog collections. We've had some success with getting people to agree to CC licenses. Limited resources for development and support of repository, digital collection management system, etc.: we do not have as much programming support as we would like, for example. We do the best we can with what we have, maintaining a commitment to standards and trying to preserve what is essential about both context and content. Arrangement and description: image filenames assigned by creators (or their cameras) can be meaningless and/or misleading, particularly when they are presented to the public as the identifier for ordering a copy of the image. In cases like this, where the original filenames were not essential or meaningful, we have renamed files. We have not yet received a collection that contains a significant amount of textual digital content; we anticipate different but substantial issues when we get to that frontier.

Copyright: we have required users to log in in order to access materials subject to copyright protection. User experience: presently, there are a number of access methods and systems in place that are not integrated. Users need to move between the disparate systems and understand how to use each system in order to encounter different parts of a single collection. We seek a solution that will be an easier way to integrate the experience into one interface. Managing access levels: we would like more access control granularity than we presently have. Rather than having collections accessible to all of the world, or all of those on the campus network or with a login. We have not yet begun discussions about how to address this issue.

Copyright. Workflow, including arrangement and description. Preservation.

Currently, materials are not in an organized database that is accessible online or easy for users to access while at the library, so basically materials that are born-digital are not really accessible at all.

Describing the large volume of born-digital materials in a scalable and practical way. To address this challenge, we are exploring the possibility of adding high level or accession level "preliminary descriptions" to the online catalog, before the records have been processed by archival staff. This applies to both born-digital records as well as traditional records. We need a better understanding of how to effectively index, search, and render results for the large volume (petabytes) of born-digital and digitized records, for a large variety of file formats. We are aiming to address this challenge by seeking assistance from a search expert to help us optimize search and display of electronic records in our online catalog. Reviewing the large volume of born-digital materials for access or use restrictions prior to making them available. The current review model and workflow is not scalable or sustainable. This challenge has not yet been addressed.

Determining how to collection and to manage born-digital materials. Determining staff resources for management of born-digital materials. Determining funding needs and resources for management of born-digital materials.

Development of an interface for patron access. Copyright. Confidential/restricted content.

Donor restrictions requiring in-house use only: in-house provision, despite user unhappiness. Copyright: take down notices, disclaimers, risk assessment.

For personal materials, the lack of a repository. We have not adequately addressed this issue.

How to deal with materials that are under copyright or otherwise of a more restricted nature. Our current systems are intended to support an open model of preservation to access. We are beginning to address the divergent needs for archive space with limited or no access and how best to manage it. We are also evaluating our collecting policies to

ensure they align with our digital acquisitions. How do we balance the need to support access to our collections with the restrictive mandates that might be required under copyright or donor agreements?

How to present digital objects in a standardized fashion or to be able to save/render legacy and/or complex formats. We have only just begun to address these challenges on a piecemeal and ad hoc basis.

Lack of a repository suited for easy upload/ingest of born-digital materials with additional tools available for supporting preservation metadata. Staff resources/time to describe born-digital materials, prepare them for uploading into an access system, particularly regarding creation of preservation metadata and other descriptive metadata. Restrictions placed by donors or by law on access to some born-digital archives and special collections, which need to be maintained in a repository but without access for a period of time.

Limited staff.

Managing permissions for various types of users (distance researchers, classes of students, faculty, TAs, etc.), which can change, are time sensitive, need to be secure. Large files can crash our server or significantly slow down the system. Structure and display of files.

Metadata: we lack full-featured metadata creation and management systems for descriptive, rights, administrative, and structural metadata. Poor solutions in the past. We are currently in the process of adapting the new Hypatia libraries on our Fedora platform. Copyright, permissions, privacy: we are working through these issues as they arise. Scaling up our operation to accommodate born-digital archival collections and other born-digital special collections may be slowed down by the need to investigate rights status, clear rights, and do risk analyses. Software development: working in the open source Fedora environment has many advantages but does require significant local investment in software development. When possible we are leveraging others' work with Fedora, Hypatia, and Blacklight and contributing code to those projects.

One of the top challenges right now for providing access to born-digital material is the inadequacy of our current descriptive tools (EAD and MARC) and their discovery and display interfaces to deal with the nature of born-digital content. The scale of the born-digital content easily overwhelms the traditional library catalog-style digital library interface (1 record per item) and the EAD record is not created or managed in a way that can take advantage of the born-digital components either (too many items to list or link them all; the text and technical metadata for objects doesn't have a container in EAD). The Hypatia project, as part of the AIMS grant, worked on ways to build interaction between these systems so that the individual digital objects can be managed in an appropriate repository environment, but discovery of them can be integrated with description of non-digital components in a finding aid. We are still working on developing this kind of system locally. A major issue related to born-digital material is restrictions on access both due to sensitive or private information and intellectual property rights. They are related, but slightly different issues. Sensitive and private information is restricted from all view unless permission is granted. In order to provide this information then we would need to grant access to some users but not to others. In some cases this would be access granted to a class of user (university affiliates), but in other cases it would be on a case-by-case basis. Issues then would be being able to identify and remove, redact, or restrict the appropriate content (not always easy to do) and to grant access to appropriate individuals. Eventually, we would like for all content to be managed through a digital repository, so that will mean that we will need some sophisticated authentication controls. Issues related to intellectual property arise due to the fact that we do not own copyright to the majority of the material we collect. When we just provided access to paper copies that were difficult to reproduce in the reading room to a single user at a time, this access was considered well within fair use. If we were to simply make access available to born-digital content online to anyone we would have dramatically changed the situation: now copying is easy and multiple people can see the content. This would increase our risk of overreaching fair use. As a way to avoid this, we will make some digital materials available only in our reading room on a dedicated computer used only for viewing content, not copying or taking notes. A third challenge is

related again to the software formats of much of the born-digital content. We suspect that most users will want to use a modern file format, especially for materials that we can make widely available on the web. In addition, we will not be able to find and support every software needed to view every file type in our own reading room for those materials only available there. For some formats, we know we can migrate to an acceptable access format (modern PDFs can be derived from early Microsoft Word formats), but for others there is no clear migration path. In addition, there will be some number of researchers who will want access to the original formats. This would mean that we would need to be able to get reasonably quick access to both a normalized access format as well as the original.

One significant challenge continues to be donor restrictions and copyright issues. For now, we are only providing access to the material within the reading room. We also would like to add a more advanced set of tools for researchers to use while interacting with the born-digital collections. Our current approach is to solicit feedback from researchers and develop plans for future tool development.

Organizing and describing research data was a significant challenge, but we believe we have developed a methodology that uses events to describe research context and RDF relationships to link resources within a research project together. We have also developed a research data application profile that enables us to provide core descriptions of research data to enable interdisciplinary reuse of data. We are working to resolve the challenge of collection level description, using EAD, within the RUcore repository. We are developing a context object methodology that uses relationship metadata to link resources and that creates generic "core" metadata at the object level, so that individual objects do not need to be described individually. We do not currently have a methodology for describing and managing websites, but are hoping that the EAD methodology, which supports hierarchical relationships, can be used to manage the more matrix-like site maps of websites.

Our ingest and storage system is brand new and partially still in the planning stage; discovery and access tools are still in development. The library is presently building a Hydra/Fedora institutional repository structure and access to born-digital records in a primary use case.

Overall processing workflows and workloads: this is related to the earlier challenge of insufficient staffing. In the short-term, we are incorporating these responsibilities along with the other primary responsibilities of existing staff. Arrangement and description: this is related to the earlier challenge of insufficient staffing. In the short-term, we are incorporating these responsibilities along with the other primary responsibilities of existing staff. Access for reference service: this is related to the earlier challenge of insufficient staffing. In the short-term, we are incorporating these responsibilities along with the other primary responsibilities of existing staff.

Processing efficiently. For physical materials, we are adopting efficient processing procedures where we organize, appraise, describe, and house materials in less and less granular ways. However, for the first few born-digital collections we processed, we found that we had to work at the item or file level. This may be because the files were from floppy disks with no discernible original order or series. An archivist had to open up, evaluate, and provide a descriptive title for every file/item. She also reviewed the material for confidentiality issues. This level of processing is not sustainable, and we are actively looking for other methods to automate this work. Many of our born-digital collections are faculty papers and contain the same sorts of files that physical faculty papers contain: collected articles authored by other individuals, letters of recommendation for colleagues, drafts of unpublished books, etc. If we provided access to these freely on the web, we might be in violation of copyright, or we might violate individuals' reasonable expectation for privacy, or we might hamper the family's ability to publish works posthumously. We developed the concept of a Virtual Reading Room so that we could provide remote access to this content online in the same way as we do to physical items in our physical reading room. While the metadata for the material is exposed publicly, you have to enter the Virtual Reading Room in order to view the content of the files. The full-text is not indexed in Google either, thus protecting the individuals about whom correspondence is written. To enter, we require that researchers complete the same application we would have them complete if they came into our physical reading room. They sign off on a copyright statement as a condition of

using the material. The Virtual Reading Room is providing a layer of risk mitigation by doing three things: 1) It shows that our intent is to provide access for educational, personal, or research purposes only, just like we have always done for similar, analog materials in our reading room. 2) It makes use of the material conditional upon users agreeing to only use the material for educational, personal, or research purposes. 3) It shifts some of the accountability for violating fair use to the user.

Providing interfaces to allow users to set access controls, and managing access to non-public items in various systems. The variety of formats and complex objects that require special user interface programming. Displaying complex objects in search results.

Restrictions: our digital asset management system is designed to be used by all units on campus, not just Libraries/ Archives. Restrictions will have to be able to be applied at unit, user, item, and collection levels. Libraries IT staff is currently working on enabling this function of the system. Archival context: while our digital content resides in our digital asset management system, archival descriptions are located in ICA-AtoM. Libraries is planning on linking the two systems together to demonstrate the provenancial and archival context of the digital items. Metadata creation: all metadata linked to digital content is created by contract digitization technicians. While the Libraries has been successful in acquiring funding for this very important task, there is no baseline funding in the Libraries budget for this type of metadata creation.

Sensitive data: we have yet to work out issues surrounding born-digital institutional records with restricted access, e.g., promotion & tenure files, president's office files, etc. Graduate school policy resists open access to electronic theses and dissertations in the institutional repository.

The basic issue of how to provide access, particularly for things that are copied to our server since that is not publicly accessible. Address is by providing access on CD or flash drive, but that requires much advance work before the patron comes. For the few collections we have that have digital content, determining and then conveying to patrons now through description what we have in analog format vs. digital format, or in both analog and digital. In practice, we provide varying levels of description about the digital content and do not always go to the extent to determine what we have in digital form vs. analog. Easily generating file lists: WordPerfect used to do this easily, Word does not.

There are concerns about our ability to provide copies of our digital content without violating copyright. We are currently reviewing existing donor agreements to evaluate what rights we have. Discovery of our digital materials is poor. EAD does not lend itself well to describing the digital materials and we do not have a metadata browse/search/discovery tool that permits a combination of item-level and aggregate description. There are concerns regarding our ability to provide original files v. derivative versions (due to redaction and/or migration) and informing potential users about both/either of the options.

Theses and dissertations: copyright issues. This is partly addressed by providing embargo options for our theses and dissertations.

Training for staff on how to arrange and describe born-digital materials. We are working to bring appropriate training to the library and to our region. Managing collections with personal identifiable information. We are trying to determine what the best strategy is for providing access to these materials and don't have a good answer. Managing user rights for restricted collections. We are investigating this as part of our Rosetta pilot project.

Unclear copyright status; verifying copyright status!!! Metadata arrangements need more standardization. How to effectively integrate into basic discovery systems.

We continue to look at methods of display for special materials (such as newer file formats, book reader (page turner) and also incorporating other metadata to support these materials (TEI, MODS). Copyright: certain materials in the institutional repository are only available with university credentials due to copyright issues.

We face the challenge of separating out confidential material. Digital access is not currently provided to collections containing this content.

We have no policies or mechanisms for delivering born-digital content at this time.

Without a Digital Asset Management System (DAMS) in place we have not been able to provide access/discovery. A current system-wide initiative is working on a solution to this. Rights management issues and a clear understanding of what materials can be made available have hindered access. System-wide initiative is currently working on a solution.

Working out issues relating to security/privacy (as mentioned above).

USE POLICY

23. Are the born-digital materials your library offers or plans to offer available to all library users or is their use restricted to certain categories of users? N=63

Some born-digital materials are available to all users; other materials are restricted	52	83%
All born-digital materials are available to all users	8	13%
Use is restricted to certain categories of users	3	5%

Please briefly describe any restrictions on the use of born-digital materials (e.g., user category, institutional affiliation, internal policy such as restriction on personnel, student-related records, etc.) N=52

Some born-digital materials are available to all users; other materials are restricted

Access restrictions may be stipulated in donor agreements, for special collections materials. Provincial privacy legislation restricts access to university archives containing personal information. The university's IP policy allows broader use and reproduction of some copyrighted materials by the university community than by external users.

As mentioned previously, our system will enable born-digital materials to be restricted by unit, user, collection, and item.

As per our normal practices, restrictions may apply if required by the donor and/or privacy laws.

At the moment, largely donor-imposed.

Based on already existing policies for restrictions regarding institutional and organization records.

Certain records are restricted per university records access policy, regardless of format.

Confidentiality.

Deed of gift restricts some materials.

Depends on the content. Items can be restricted to individuals or to specific IP ranges.

Digital materials may be subject to the same sorts of restrictions as analog materials: law (FERPA, etc.), university records policies, and donor-imposed restrictions. There are no restrictions unique to digital materials.

FERPA or other privacy laws, copyright, IPR, internal hierarchical operations, donor restrictions.

FERPA-based as well as institutional policy.

Institution affiliation for electronic theses and dissertations.

Just like any other collection that might have restrictions.

Licenses or agreement restrictions.

Like any material in our collection, some will have restrictions on access, determined by the deed of gift. Otherwise, all content will be available to anyone, although material for which we do not own copyright will most likely only be available by physically visiting the reading room.

Materials may be restricted based on donor agreements, state records laws, privacy considerations, and university policies and regulations.

Materials where we hold copyright and not subject to restrictions (primarily university records content previously released to the public & web-content) are available online. Other materials, where there is uncertainty regarding copyright, are accessible locally. Still other materials are restricted as per donor agreements or to protect SEI.

Materials that we own the copyright to we distribute freely. Other items might be made available to researchers for purposes of private research and personal use after signing a researcher agreement with the university.

Only user restrictions would be copyright and need to clear use with owner of images, as well as any security classified (these are both access and use restricted) restrictions that may pertain to certain images in an accession.

Restrictions are based on user category. Some restricted institutional materials may only be open to certain categories of users from the university.

Restrictions based on federal law, donor requests, type of record.

Restrictions based on information content, just like all other archives.

Restrictions can be for reasons of copyright, contract law, or privacy. Different collections or material types raise different issues. Our informal categories of access are: world-readable, on-campus access, on-site access, access by permission for research, and access only by content owners or curatorial staff.

Restrictions might include issues related to access or copyright. Personal data would be further restricted as would any donor-level agreements.

Restrictions on personnel and student records.

Restrictions on some born-digital materials include: lack of copyright to make the content available freely online; donor restriction for a period of time due to sensitivity of information in the materials; legal restrictions to content, such as documents containing personally identifiable information relating to medical treatment, etc.

Restrictions on student-related information protected by FERPA. Restrictions on embargoed material.

Some faculty data sets deposited to the digital repository have restrictions.

Some items restricted to on campus only.

Some materials are restricted to university-only access because they are student-related records, have copyright restrictions, etc. Some records held by the Medical Center Archives contain PHI and are protected under the HIPAA Privacy Rule and the HIPAA Security Rule.

Some materials may be restricted as a result of policies and/or regulations, or donor agreements.

Some materials *may* be restricted to non-community users, perhaps. As of now, everything is open to everyone and we will try to keep that status as long as we can.

Some material may be restricted to on-campus use only (i.e., in library or reading room) or to members of the campus community (i.e., faculty, staff, students).

The dark archive is accessible to library staff only. Born-digital materials have the same access restrictions as their analog predecessors. Graduate school policy resists open access to ETDs. Researchers deny open access to data collections.

The format of the records has not changed who we make records available to. Our standard user categories (i.e., university staff, researchers, etc.) still apply.

The restrictions would be the same as we have for paper records—restrict personnel or student-related data, donor restrictions.

The same restrictions that apply to paper documents will apply to born-digital materials. We have a five-page access policy governing these issues that I cannot restate here.

There are internal policies relating to personnel, such as university personnel records containing personal information.

University records are restricted for a set time period and then open to the public. However, the creating office has access to the records during the restriction period. The restrictions vary depending on the nature of the creating office. We also have some collections that are restricted based on donor request.

University records are restricted to certain staff. Some manuscript collections carry restrictions to certain categories of users. However, most collections are open to all users.

User category.

User category and institutional affiliation.

Varying restriction periods on certain types of records (institutional records, personnel, student) and potential other restrictions per individual gift agreements.

We embargo some resources, such as ETDs and research data, at the request of the student or faculty member. These embargoes generally represent a desire to finish a research project, publish a book or article on the findings, etc. We will also restrict parts of a research project indefinitely if the data has privacy issues.

All born-digital materials are available to all users

All content is available to all users; however, not all content is available on the web.

As a public institution, institutional records are available for public scrutiny within certain exceptions outlined in law. Any other restrictions would be dictated by donor agreements. There are no uniform restrictions, and we have such a minimal amount of born-digital materials right now there are not any identifiable categories.

Copyright and university policy are the main restrictions for the content we have collected to date.

Use is restricted to certain categories of users

For current processed content, only MARBL users who have registered with MARBL and have access to the researcher workstation in the reading room can use born-digital materials.

Restrictions based on internal policy.

Some of our born-digital collections are available only in a Virtual Reading Room. To gain access to our Virtual Reading Room, all researchers must submit an application to use a collection and agree to follow our rules of use. So long as researchers do this, they may use the materials.

24. Does your library require users to complete any registration process before using born-digital materials? N=60

Yes	25	42%
No	35	58%

If registration is required, please briefly describe the process. N=34

Answered Yes

All on-site users register at first visit, regardless of what materials they are using.

All researchers who use MARBL content must first register. Researchers fill out a questionnaire, show identification, and have a brief orientation/interview with Research Services staff.

All special collections users complete the same registration process (using Aeon system). No additional registration, unless required for a particular collection.

All users who visit our reading room are required to register by showing a valid form of ID and filling out a registration form (users who are affiliated with the university are pre-registered through our university's authentication system). Since many materials will be available only in the reading room, users will need to register before use. We are exploring the idea of making other materials available freely to university users from any location through a login, or making material available through campus networks, but those uses have not been fully worked out and may not be viable or needed.

Currently, it still is a paper process. We are developing a "click through" approach in the new repository environment.

Currently, we use the same registration process required for physical manuscript collections. The user fills out the appropriate paperwork and then has an interview with one of our curators. This has been fairly easy to enforce because the born-digital materials currently held in special collections are used in our reading room. There is no equivalent registration process for the born-digital materials in our digital collections, institutional repository, or open journal system. We have other born-digital materials that may not be accessed without the permission of the principal investigator of the research project that the materials pertain to.

Fill out contract outlining what you want access to. Researcher assumes liability should they republish or distribute the material. Once signed form is returned or payment received (cost of digitizing an analog recording, photograph, etc.) we provide the copies or access.

If born-digital materials contain sensitive data, researchers may be required to apply to the Institutional Review Board to gain access.

In the Rare Book and Manuscript Library, registration is done through standard Rare Book registration procedure. Elsewhere within the library system it is via informal request to curatorial staff or administrators. This will continue to vary depending on the type of material or collection, its nature, and its custodial location.

Materials available online via DSpace or Archive-It do not require user registration. All other materials require a user to register on-site as per our default researcher policy.

Normal archives patron registration process.

Normal registration process.

Patron registration form.

Reader registration and being on-site at the library is required to access all restricted collections.

Registration through paper form is required of all users. Special Collections and Archives staff enter the information into a restricted database and shred the paper forms.

Required to fill out online registration form and note which collection they are accessing.

Same as for all SC/A materials.

Same as for on-site use of any special collections.

Since all access is on-site (except for some oral histories), they would undergo the same reader registration process required for users of paper collections. Processed oral histories that were originally recorded digitally (not digitized from analog tape) are available without registration in the Louisiana Digital Library, a CONTENTdm site.

The same as we would for paper materials.

They complete the standard Patron Use form that all users complete.

Users fill out a form. After completing it, staff review it and allow researchers to create an account in our DSpace system to access the material. The registration applies only to born-digital materials that we put into our Virtual Reading Room, due to copyright or confidentiality issues. Some of our born-digital material is open to all users online.

We are rarely able to make born-digital materials accessible, but when we do, patrons must complete the same registration process as patrons using papers in our reading room.

While we cannot provide access at this time, access to born-digital records for onsite users will require registration.

Yes, the same as any other user.

Answered No

Haven't thought about that.

However, on-site researchers have to register into the research rooms and on-site researchers using a library-provided computer have to register use of that computer, but not to access born-digital materials per se.

Not for materials in the IR or on the open web, but if made available only within special collections, the same registration process that applies to all users would be used. A paper form is filled out, and the researcher must provide photo ID.

Traditional methods of registration include library accounts and registration upon entering the archives. Online methods include IP range restriction, Shibboleth authentication, and tracking from web analytics such as Google Analytics.

We don't, but this is an interesting thought, and we may.

Will depend on the materials.

With online access no, but if one is using the object in the reading room, there is a registration process.

Other Comments

It would depend. Not for our LUNA collections, but if we allowed in-reading room use, yes.

We will be exploring this issue at a later time and cannot provide a response until our exploration of this topic concludes.

ADDITIONAL COMMENTS

25. Please submit any additional information about processing and managing born-digital materials at your institution that may assist the authors in accurately analyzing the results of this survey. N=20

As much as possible we treat all archival content the same way in terms of policies and procedures. For born-digital the main differences are technological issues that are mainly internal and do not affect patron policies.

For university institutional records, the Records Manager will be heavily influential in acquiring born-digital materials that are authentic and reliable by working with creators before records are created and will have to work closely with Archives staff in ensuring their authenticity and reliability are preserved during their transfer to archival custody.

In early stages of managing born-digital materials beyond basic content such as e-dissertations and theses. Currently assessing future directions for growing born-digital collections, including many of the questions raised in this survey.

Libraries and Archives both report to CIO. Digital curation approached differently due to different missions, but Libraries and Archives collaborate where we can.

Other than ingest, access, and preservation issues, we don't treat born-digital materials any differently than we do paper materials. We intend to apply the same policies and procedures to born-digital materials wherever possible.

Policies and practices differ across special collections units within the library, although our collaboration is increasing as we seek to find shared solutions. Variances in practice are clarified in comments throughout the survey.

Processing and management of born-digital materials in the library and across the university has been somewhat fragmented, developing within functional "silos" over time. Current campus-wide information systems planning initiatives and also strategic planning within the library will reduce this fragmentation of effort and facilitate future management and larger-scale ingest of born-digital special collections and university archives materials. One important aspect of these initiatives will probably be the development of digital repositories that can be used by different groups within the university with similar storage and access needs.

Separate from Special Collections, the institutional repository ingests research data, non-commercial e-only publications, and electronic theses and dissertations.

The processing and management of born-digital materials is currently done on an ad hoc basis. We are working to create procedures and policies to institutionalize our practice but there is little literature to use as a basis for this.

This survey is very timely as we have had two recent potential donors want to give born-digital material to the South Carolina Political Collections (SCPC) archive and the special collections librarians have begun talking about these very issues. We are in the early stages of creating policies and strategies for preservation and access, but we know we need to. SCPC has already acquired a fair amount of legacy media and electronic files, but the South Caroliniana Library and Rare Books are not far behind in collecting born-digital materials as well.

We are acquiring digital content, but not in great volumes: a few primarily analog collections have come in with floppy disks, CDs, etc., we have received a few born-digital photograph collections, we preserve some born-digital university records (including photographs) and community publications, we work with ETDs, and we have preserved some community-related web content via Archive-It.

We are at the early stage of development in terms of process flow and content management. Very little has been operationalized. Different types of born-digital material (e.g., e-archives, web archives, research data, audio and video oral histories) have different requirements, staffing needs, timetables, etc., and will necessarily have different workflows and ingest routes.

We are currently building the IT staff capacity to help support the digital library/archives initiatives.

We are currently in a transition phase. We have newly created positions and new hires (e.g., Digital Archivist) brought in to more pointedly better address issues/challenges of collecting, managing, preserving, and delivering born-digital and digitized content.

We are farther ahead working with materials in the digital repository than working with archives and special collections. Archives and Special Collections was not part of the digital repository development or the selection of materials to place there until very recently.

We are in the process of moving management of our digital collections as a series of separate projects to a program-based coordinated approach across the institution. Institutional archives are collected and managed through the Office of the Librarian.

We are still in the very early stages of determining how to manage born-digital materials. This survey has provided much food for thought.

We have begun planning for everything mentioned in this survey, but have really only begun implementation. I've answered the questions based on what our plans are, but in many cases we have not actually completed implementing these plans.

RESPONDING INSTITUTIONS

University of Arizona
Arizona State University
Brigham Young University
University of British Columbia
Brown University
University of Calgary
University of California, Irvine
University of California, Riverside
University of California, San Diego
Case Western Reserve University
University of Chicago
University of Colorado at Boulder
Columbia University
University of Connecticut
Cornell University
Dartmouth College
Duke University
Emory University
University of Florida
George Washington University
Georgia Institute of Technology
University of Illinois at Urbana-Champaign
Iowa State University
Johns Hopkins University
University of Kansas
Kent State University
University of Kentucky
Library of Congress
Louisiana State University
University of Louisville
University of Manitoba
University of Massachusetts, Amherst
Massachusetts Institute of Technology
Michigan State University
University of Minnesota
University of Missouri
Université de Montréal
National Archives and Records Administration
University of Nebraska–Lincoln
University of New Mexico
University of North Carolina at Chapel Hill
North Carolina State University
Northwestern University
University of Notre Dame
Ohio University
Ohio State University
Oklahoma State University
Pennsylvania State University
Purdue University
Rutgers University
University of South Carolina
Southern Illinois University Carbondale
University at Buffalo, SUNY
Syracuse University
Temple University
University of Tennessee
Texas Tech University
University of Virginia
Washington State University
Washington University in St. Louis
University of Waterloo
University of Western Ontario
Yale University
York University