2017-07-03

Disaster Planning and Trustworthy Digital Repositories

Frank, Rebecca D.

Disaster Planning and Trustworthy Digital Repositories

Rebecca D. Frank
April 20, 2012

1. Introduction

Disaster response and recovery planning remains one of the most important components of a preservation program in digital repositories, and also one of the least understood. The adoption of standards and models for preservation such as the Audit and Certification of Trustworthy Digital Repositories and the Open Archival Information System (OAIS) model have helped to clarify and illuminate best practices in the digital preservation community. However, our understanding of disaster planning for digital repositories remains limited.

In an article written for Wired Magazine, Chris Anderson argued that we are currently in the "Petabyte Age" (Anderson, 2008). This age is marked by an exponential increase in digital data. This proliferation includes scholarly research data as well as digital information created for entertainment and personal use, "the digital universe — information that is either created, captured, or replicated in digital form — was 281 exabytes in 2007. In 2011, the amount of digital information produced in the year should equal nearly 1,800 exabytes, or 10 times that produced in 2006" (Gantz et al., 2011, 3).

In terms of storage, "2007 marked the 'crossover' year in which more digital data was created than there is data storage to host it" (Berman, 2008, 52). This tipping point, the point at which data created outpaced our capacity to store data, is significant for the digital preservation community. It is at this point when decision making for digital preservation must focus not only on how to preserve data, but also on what to preserve.

These decisions are based on any number of criteria, but the important factor to consider for digital preservation and disaster planning is that the information selected for preservation in digital repositories has ultimately been selected because of its value. "While the costs of maintaining digital preservation capacity are not insignificant, the costs of the alternative are often greater. Re-creating research data sets can be prohibitively expensive; in the extreme, it may be impossible to re-create lost data"(Beagrie, Chruszcz, & Lavoie, 2008, 16). The importance and uniqueness of data such as this, compounded with the difficulty or impossibility of recreating lost data, makes a strong case for preservation. Because of this need to preserve the data that is held in digital repositories, disaster planning is a particularly important activity. The digital preservation community is developing an awareness and understanding of the concept of disaster planning as part of a digital preservation program, but a thorough understanding of disaster planning in practice has not yet been achieved.

The goal of this study is to understand if digital repositories that have a preservation mandate are engaging in disaster planning activities, particularly to further their pursuit of trusted digital repository status. In cases where digital repositories are engaging with disaster planning, the study also examines the process of creating disaster response and recovery plans, with a focus on how these activities are integrated into the management of the digital repositories.

This study focuses on the practices of digital repositories that have either sought trusted repository status, have undergone some type of self-audit, or have expressed a commitment to pursuing this type of certification process in the future. As the literature indicates, disaster planning is generally understood to be part of the requirements for trusted repository status, but the details of such planning activities are not well documented or understood.

2. Literature Review

2.1 Digital Preservation

In order to understand disaster planning for digital repositories, it is important to first examine digital preservation and the relationship of preservation to disaster planning. Disaster planning for digital repositories has the same intellectual roots as digital preservation, "digital preservation can encompass a range of activities, from simple replication and storage to more complex transformation, depending on the assessed value and risk to the target content" (Hitchcock, Brody, Hey, & Carr, 2007, 1). Francine Berman states that preservation actions are, "actions undertaken to ensure the long-term viability and availability of the authoritative nature of digital material. Preservation actions should ensure the material remains authentic, reliable, and usable while its integrity is maintained; such actions include validation, assigning preservation metadata, assigning representation information, and ensuring acceptable data structures and file formats" (Berman, 2008, 55). In short, digital preservation consists of those actions that ensure the viability and authenticity of digital objects over time and disaster planning is one of those actions.

Disaster planning or preparedness in a traditional sense "refers to a state or situation of the libraries in which they are well prepared to prevent severe library damage from potential disasters" (Wong & Green, 2006, 72). And more specifically, a disaster plan is a document that describes policies and procedures which have been created to prevent, prepare for, respond to, and recover from a disaster (Muir & Shenton, 2002). Analogous to analog collections, disaster planning is an essential activity for digital repositories (Patkus & Motylewski, 1993). The concepts underlying disaster planning for analog materials can be applied to digital repositories in that the long-term preservation of digital materials depends on the ability of an organization to prevent, prepare for, respond to, and recover from disaster events.

In 2007, the Center for Research Libraries, The Digital Curation Center, DigitalPreservationEurope, and NESTOR met and identified a list of ten characteristics of digital preservation repositories (Center for Research Libraries [CRL], 2007). This list "provides a structure that informs the processes and outcomes" of repository audit and certification processes such as the Trusted Repository Audit and Certification (TRAC), which will be discussed in greater detail below (McHugh, 2008, 133). The characteristics are:

1. The repository commits to continuing maintenance of digital objects for identified community/communities.

2. Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
8. Fulfills requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

These criteria relate both directly and indirectly to disaster planning. Specifically, criteria regarding maintenance (1, 3, 6, and 7) assume that the repository will be able to maintain digital objects and their metadata over time, presumably in spite of any disasters that may occur. Criteria regarding preservation and security (9 and 10) are also significant for disaster planning in that disaster planning efforts are meant to ensure long term preservation and security of digital objects.

One key problem facing the field of digital preservation is the sheer volume of data. While this may not be a problem at the individual repository level, as each repository is able to accept for deposit only that data which meet their specified criteria, it is a problem for the community as a whole, "the scale of digital creation is far outpacing the capacity to store the data" (Berman et al., 2010, 9). This problem of scale has been widely documented (e.g. Berman, 2008; Hey, 2003). And it is from this problem that others arise. Specifically, problems concerning how to ensure the long-term viability of sustainable digital repositories while continuing to grow. This problem also leads to different proposals for disaster mitigation solutions, two of which are described below.

2.1.1 LOCKSS

Some approaches to digital preservation have implicit disaster planning strategies built in. One such approach to digital preservation is the LOCKSS (Lots of Copies Keeps Stuff Safe) system. LOCKSS is modeled on the system used by libraries to preserve physical content through duplication of resources across multiple distributed organizations, "the phrase 'distributed digital preservation federations' is being used increasingly to describe cooperatives of geographically-dispersed institutions who are banding together to form solutions to the digital preservation problem" (McDonald & Walters, 2010, 1). For digital preservation, "a combination of massive replication, rate limitation, inherent intrusion detection and costly operations can produce a peer-to-peer system with remarkable ability to resist attacks by some extraordinarily powerful adversaries over decades. Its lack of dependence on long-term secrets and stable

identities blocks many of the paths by which systems are typically attacked" (Maniatis et al., 2005, 42). This particular system of preservation is reliable because it allows multiple repositories to share responsibility for digital objects. While each partner in a LOCKSS system is indeed responsible for maintaining their copy of the items, they are able to restore any and/or all of the items in their repository from another partner in the event of data loss.

While it is often not directly stated, the 'lots of copies' part of a LOCKSS system is, in effect, meant to preserve the data that may suffer a disaster at one location by providing duplicates across several locations. Articles such as those by Maniatis et al. (2005) highlight the strength of a LOCKSS network to resist "attack" and "random storage faults," both of which can be considered disaster events (30). In a study published in 2007, Schroeder and Gibson (2007) conducted a survey of "field-gathered disk replacement data from a number of large production systems, including high-performance computing sites and internet services sites. About 100,000 disks are covered by this data, some for an entire lifetime of five years" (1). The study found that "in the field, annual disk replacement rates typically exceed 1%, with 2-4% common and up to 13% observed on some systems," a failure rate that was significantly higher than the authors expected (Schroeder & Gibson, 2007, 1). This suggests that the random storage faults discussed by Maniatis et al. are indeed likely to occur. However, these articles do not focus specifically on disaster response and recovery planning, rather, implying that the duplication for long-term preservation will allow the system to overcome any type of disruption or loss in service.

2.1.2 iRODS

Another approach to digital preservation is the Integrated Rule Oriented Data Systems (iRODS) software that has been developed by the Data Intensive Cyber Environments group (DICE). iRODS is "a second generation data grid system that facilitates data management spanning large geographic areas and across administrative domains" (Data Intensive Cyber Environments Group [DICE], 2008, 1). As described by Moore, "the iRODS data grid is a generic data management infrastructure that can be tuned to support data preservation, data publication, data sharing, or data analysis through specification of appropriate data management policies" (Moore, 2008, 73). The iRODS system is not specifically a preservation system, but it can be used to facilitate and support preservation by mitigating against risk. "When a user or organization stores data with associated metadata in a data grid, they apply policies to ensure that the resulting collection will meet their goals. Such policies include disaster recovery (syntactic replication), [and] persistent preservation for the long term (temporal replication)" (DICE, 2008, 2). In other words, iRODS is a system that allows repositories to create and enforce rules and policies, therefore ensuring consistency within the repository (Rajasekar, 2010). These rules and policies include those relating to long-term preservation.

2.2 Digital Curation

Digital curation is a value proposition. According to Walters and Skinner, "digital curation refers to the actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content" (Walters & Skinner, 2011, 5). Similarly, Maureen Pennock at the Digital Curation Centre describes digital curation in the

following way, “digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for both current and future use: in other words, it is the active management and appraisal of digital information over its entire life cycle” (Pennock, 2007, 1).

In Pennock’s view, curation is different from preservation in that preservation has a more narrow focus on maintaining continued access to digital materials over a long span of time. Proponents of digital curation argue that digital curation can take place with collections that are the subject of digital preservation efforts, and that it can also take place with collections that are not meant for long-term preservation. Others in the field of digital preservation do not necessarily agree. Given these differences, the question of whether disaster planning falls into the category of digital curation arises. While disaster planning is not featured prominently in the digital curation literature, it could be argued that disaster planning is indeed an important activity for digital curation. While an item or collection is needed, the repository must be prepared for any disaster that threatens the value of the information. Despite the fact that digital curation does not require long-term preservation, the process of full lifecycle management of digital resources means that those resources must be managed and preserved for as long as they are needed. Susceptibility to disasters is a problem not only if it interrupts access to collections but also if it threatens the integrity of those collections, whether they are needed for one year or twenty.

2.3 Trust

Another important element of preservation and disaster preparedness for digital repositories exists at the repository level, and that is the concept of trust. Garrett and Waters make the claim that, “for assuring the longevity of information, perhaps the most important role in the operation of a digital archives is managing the identity, integrity and quality of the archives itself as a trusted source of the cultural record. Users of archived information in electronic form and of archival services relating to that information need to have assurance that a digital archives is what it says that it is and that the information stored there is safe for the long term” (Garrett & Waters, 1996, 23). The implication here is that if a repository is not trusted by users, then the data stored in that repository is not preserved. Users must be able to trust that the data contained within a digital repository is what it purports to be, and one of the ways that users judge integrity of digital objects is through trust in the repository.

Trust is also an important component of disaster planning in that one way in which repositories gain trust is through demonstration of preparedness. Repositories demonstrate their ability to preserve their content through disasters by making disaster planning documentation available to the community, by conducting self-audits of best practices and making the results available to the community, or by undertaking a process of audit and certification as administered by an external organization.

2.3.1 TRAC

The concept of trust has emerged as a community standard for digital repositories; specifically, the assignment of Trusted Repository status through certification. Three examples of which are

the Data Seal of Approval (http://datasealofapproval.org/) which originated in the Netherlands, DRAMBORA (http://www.repositoryaudit.eu/) which was developed jointly by the Digital Curation Centre and DigitalPreservation Europe, and Trusted Repositories: Audit and Certification (TRAC) (http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0) which is administered by the Center for Research Libraries (CRL) in the United States. TRAC certification is based on the Trustworthy Repositories Audit & Certification: Criteria and Checklist (ISO 16363, 2012). Each of these certifications require that the repository seeking certification undergo an audit process, although the process of TRAC certification is more rigorous and time consuming than Data Seal of Approval certification, and DRAMBORA was designed to be a self-audit process (e.g. McHugh, 2008; Sesink, 2010).

Disaster planning is a core construct of the TRAC audit and certification requirements. Sections 5.1 and 5.2 are most explicit:

> **"5.2.4 The repository shall have suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s)**" (ISO 16363, 2012, 78).

Disaster planning is an explicit element of the TRAC certification process, and is an implied (but not directly stated) element of the Data Seal of Approval certification process. While the guidelines for TRAC certification do not provide detailed instructions or requirements for disaster planning, the certification does require that the repository be able to demonstrate disaster preparedness. This disaster preparedness is generally demonstrated through the creation of a disaster plan or, more accurately, a suite of disaster planning documents. The checklist states that, "the repository shall identify and manage the risks to its preservation operation and goals associated with system infrastructure" (ISO 16363, 2012, 65).

Of the three repositories included in this study that are TRAC certified, two created their disaster planning documentation for the audit (Portico and Chronopolis), and the third (HathiTrust) completed the audit without disaster planning documentation in place. HathiTrust has committed to completing their disaster planning documentation before their next audit.

2.3.2 DRAMBORA

The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) is another method for assessment of digital repositories "developed jointly by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE)" (McHugh, 2008, 131). DRAMBORA assessment "requires repositories to expose their organization, policies and infrastructures to rigorous scrutiny through a series of highly structured exercises, enabling them to build a comprehensive registry of their most pertinent risks, arranged into a structure that facilitates effective management" (McHugh, 2008, 131).

The focus on risk in the DRAMBORA assessment can arguably be seen as analogous to the TRAC requirement for disaster preparedness. The DRAMBORA assessment, in fact, has a stronger

focus on risk (disaster) management and mitigation as the entire assessment is based on a repository's ability to manage and respond to risks. In this respect, "DRAMBORA represents a bottom-up approach that takes risk and risk management as its principle means for determining digital repositories' success and for charting their improvement" (Innocenti & Vullo, 2009, 139).

DRAMBORA uses language that places a heavy emphasis on risk management, in service of evaluating the preservation efforts of repositories, "risk is utilised as a convenient means for comprehending repository success - those repositories most capable of demonstrating the adequacy of their risk management are those that can have, and engender, greater confidence in the adequacy of their efforts. Preservation is after all, at its very heart, a risk management process. The fundamental temporal challenges of preservation are naturally complicated by future uncertainties" (Innocenti & Vullo, 2009, 144).

While the TRAC certification process discussed disaster preparedness in only one section of the audit documentation, "the DRAMBORA process focuses on risks, and their classification and evaluation according to individual repositories' activities, assets and contextual constraints" (Innocenti & Vullo, 2009, 141). The result of this process is "a determination of the repository's ability to contain and avoid the risks that threaten its ability to receive, curate and provide access to authentic and contextually, syntactically and semantically understandable digital information" (Innocenti & Vullo, 2009, 141).

Unlike TRAC and DSA, the results of DRAMBORA audits are not necessarily made public. Of the repositories included in this study, HathiTrust has acknowledged the completion of a DRAMBORA audit but the results of that report are not publicly available.

2.3.3 Data Seal of Approval

Data Seal of Approval (DSA) is an assessment consisting of sixteen guidelines, which "recognize that responsibility for archival quality data is shared amongst three groups: producers for the quality of the research data themselves, the repository for the quality of data storage and availability, and consumers for the quality of data use" (Ball, 2010, 31). Underlying these guidelines are the following five criteria, which determine whether data can be considered sustainably archived (Sesink et al., 2010, 1):

1. The research data can be found on the Internet.
2. The research data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data.
3. The research data are available in a usable format.
4. The research data are reliable.
5. The research data can be referred to.

The guidelines themselves are organized into three sections, focusing on the data producer, the data repository, and the data consumer. Guidelines four through thirteen focus specifically on

the data repository, and while disaster planning and risk management are not explicitly discussed the focus on digital archiving, long-term preservation, and lifecycle management are relevant to the area of disaster planning and risk management.

Of the repositories included in this study, ICPSR is DSA certified. The results of the certification audit are available via the DSA website, as are all of ICPSR's disaster planning documentation.

One key difference between TRAC, DRAMBORA, and DSA, despite the fact that the goal of each assessment is to determine the fitness of a repository to care for and curate collections as well as provide long-term preservation solutions, is that TRAC and DSA provide strict guidelines for performing an audit while DRAMBORA provides a framework that can be adapted to fit the needs of the repository (e.g. Ball, 2010; CRL, 2007; Patel, 2007; Sesink, 2010).

Despite these differences in philosophy and degree of formality, TRAC, DRAMBORA, and DSA all specifically include requirement for repositories to have disaster planning and risk management documentation (e.g. McHugh, 2008; Ross, 2006; Sesink, 2010).

2.4 Threats to Digital Collections

Disaster planning, disaster mitigation, and risk management activities arise from real and imagined threats to collections (e.g. Aikin, 2007; Altman et al., 2009; Anderson, 2005; Cervone, 2006; Maniatis et al., 2005). These threats can be divided into four broad categories:

> There are many threats to archived digital information. *Physical* threats result from chance, natural events, or age, and include failures in media, hardware, storage facilities, and so forth. *Technological* threats include format obsolescence and destructive software errors. *Human* threats include curatorial error, and insider and outsider attacks. *Institutional* threats include mission change, change of legal regime, or economic failure. Many of these threats are ameliorated through replication of the materials to be preserved, combined with regular auditing (Altman et al., 2009, 181).

As with disasters for traditional analog collections, digital disasters can be caused by physical, human, and institutional threats. To this list we can also add the category of technological threat. While many incidents that fall into this category could also fall into one of the other three, and in fact nearly every disaster at a digital repository will involve some sort of technology failure, this type of incident is unique to digital repositories and can in fact happen independently of the other three disaster types. Frank Cervone identifies three types of disasters for digital repositories: technical threats, natural threats, and human threats (Cervone, 2006, 175). This is similar to Altman's categorization, although perhaps less specific.

Digital repositories face threats as discussed above, but also threats that are new and unique to digital resources, "at one time, fire and water were the two great threats to a library's collection and records. Now they have been joined by other, more insidious, but just as disastrous threats: computer viruses, hackers, file format obsolescence, storage media

degradation or obsolescence, platform dependence, catastrophic system failure, natural disasters, terrorist attacks, and simple neglect" (Anderson, 2005, 9). Preservation for traditional physical collections generally involves protecting collections from active dangers, but barring a disaster the objects generally do not require regular intervention for ongoing maintenance to mitigate against "silent corruption" (Constantinescu et al., 2008, p. 108). Silent corruption occurs "when incorrect data is provided to the user, e.g., written to the memory or I/O system, and no error is triggered" (Constantinescu et al., 2008, 108).

This is not the case with digital objects. Without regular active interventions, digital objects will quickly become obsolete, "a digital preservation plan should include scheduled migration of materials to new media, offsite backup, a disaster recovery plan and scheduled regular testing of media and backups" (Anderson, 2005, 10). This threat of obsolescence is in fact a disaster as it poses a major threat to the repository (Anderson, 2005). Other disasters that threaten digital repositories include power grid events, service interruptions, and data corruption (Constantinescu et al., 2008).

2.5 Planning for Disasters

Disaster planning for digital repositories is in many respects more complicated than disaster planning for traditional collections. With the advancement of technology, and the move toward digital resources, including both born-digital items and the digitization of physical items, disaster planning has taken on new, technologically-driven and focused aspects. While general recommendations and instructions for handling damaged materials will be suitable across nearly all traditional collections, this is not the case for digital repositories. The disaster plan will necessarily reflect the policies and procedures of the organization, and these policies and procedures will be a reflection of the preservation activities in which the repository chooses to engage.

For example, a disaster plan for a repository that chooses to only back up their data up on magnetic tape will look quite different from the plan for an organization with a mirror site at a remote location. While each solution is meant to address the same threats described above, the actions required to carry out each preservation activity are quite different, and the way that data would be restored after a disaster event are also very different. Myles suggests the following activities that are generally applicable across many types of repositories, (1) Inventory all computer hardware and software. Describe what services they support; (2) Determine what services are the most critical to your library. Describe the procedures for continuing these services in a disaster situation and how the library can recover from the disaster; (3) Make sure that computer data is backed up on a regular basis. Mission-critical data should be copied and stored off-site; . . . (5) Review the list of contingency procedures to determine ways to reduce the length of service disruption (Myles, 2000, 49).

Disaster planning documents for digital repositories tend to assume that disasters, large and small, will occur and that the organization will have to recover. While a certain amount of

prevention can be helpful, "maintenance is always cheaper than recovery or re-creation, so it makes good business sense to plan for and fund preservation," there are some types of disasters that are outside of the control or influence of the repository (such as power events) (Anderson, 2005, 9). For these types of disasters, repositories must do what they can to mitigate data loss, "data loss in complex systems, whether through natural disaster or more likely through human error, is inevitable. Recovering from these phenomena is an organizational challenge that will become an ever-increasing dilemma for research, educational, and cultural organizations as their artifacts become born-digital in nature" (McDonald & Walters, 2010, 4).

In anticipation of the need to recover from data loss, repositories are moving toward the widespread adoption of best practices for preservation, "current best practice is moving toward a systematic approach to data replication, which includes maintaining consistent unique identifiers for each resource; explicit metadata describing the resources, provenance, version, and associated rights; and a managed set of replication services. Best practice is moving toward more systematic and explicit replication policies that include multiply replicating entire collections off-site, explicit versioning, and a process of regularly refreshing and verifying replicated content" (Altman et al., 2009, 181-2). These best practices also contribute to the granting of trusted repository status as described above with TRAC, DRAMBORA, and DSA certifications.

Literature discussing disaster planning for digital repositories is sparse, and as such discussion is necessarily limited. However, the general trends discussed above, and the recognition by the community that disaster planning is a beneficial and recommended action for digital repositories, is promising and suggests that this is an area that will continue to expand.

2.5.1 Disaster Response and Recovery

Literature discussing disaster response and recovery for digital collections is also sparse. While there is some literature discussing business continuity planning for private sector companies, such as financial institutions, this literature does not address some of the important and specific peculiarities of digital repositories (e.g. Andrew, 2008; "Best Practices in Disaster Recovery Business Continuity Planning," 2008; Cousins, 2007; Nollau, 2009; Wheatman et al., 2001). For example, budgetary considerations are completely different for a private company than a nonprofit digital repository that is likely part of an educational institution or library. Additionally, the type of data held in each repository may also be quite different.

Roy Tennant argues, "Once the emergency has passed, you should know what steps must be taken to get everything back up and functioning. Specifically, you should know in advance how to install new hardware and software, retrieve data from a backup system, and get everything back online" (Tennant, 2001, para. 14). This advice is true whether data is backed-up on magnetic tapes or at a mirror site, and echoes the need for disaster preparedness training for staff of digital repositories. Just as staff at traditional organizations run tabletop exercises to test out the disaster plan, staff at digital repositories should do the same. Staff at digital repositories should, in fact, go through the entire process of restoring their data from backup

so that any problems in the process can be addressed before such action is necessary.

In responding to a disaster where the information content is damaged, an important and unique consideration for digital repositories is the issue of managing damaged equipment. While damaged equipment certainly should be replaced if necessary, it is also a good idea to keep the damaged items until the system has been completely restored, "tapes previously thought of as unreadable later turned out to have useful data. Defective hard drives too may have recoverable data. Do not let anyone dispose of any equipment or data sources until the emergency is truly and completely over" (Brennan & O'Hara, 2002, 72).

Despite this apparent lack of literature regarding disaster response and recovery for digital repositories, literature relevant to disaster planning can be found in several other areas. Risk management and business continuity (or continuity of service) planning are two such areas, and both will be discussed in greater detail below.

2.5.2 Risk Management

Risk management is a term that is found in both the literature and in common discussion of disaster planning for digital repositories. It has been argued that, "protecting digital objects against threats is equivalent to reducing the risk of those threats, which is the main goal of the broad area of *Risk Management*" (Barateiro, 2010, 5). The certification and assessment programs discussed above, TRAC, DRAMBORA, and DSA are based on the concept of risk management for digital repositories. In each case, the trustworthiness of a digital repository is evaluated based on that repository's ability to manage risk and/or mitigate the effects of disaster events on the repository. The phrase 'risk management' is used in some cases to describe disaster planning activities, and a risk management approach can be used to inform disaster response and recovery planning activities. It is also true, however, that risk management literature does not place a strong emphasis on disaster planning over digital preservation in general. Rather, risk management literature tends to take a more broad view of risk management in terms of digital repositories and, as mentioned above, discuss risk management in relation to long term digital preservation activities and strategies rather than disaster planning.

2.5.3 Business Continuity Planning

As mentioned above, and similar to risk management, some of the literature regarding business continuity planning (BCP) can also be used for disaster response and recovery planning for digital repositories, "BCP is concerned with the recovery and resumption of activities across the entire organization" (Cervone, 2006, 174). As with nearly all of the literature identified in the area of digital disaster planning and digital disaster response and recovery, the article by Frank Cervone provides solid and clear advice but does not provide discussion or analysis of the disaster planning efforts of any particular organization. While academic articles provide interesting anecdotal cases, and business materials (such as those prepared by Gartner research) provide sound advice, none provide analysis of current practices (e.g. Battersby, 2005; Fletcher, 2006; Heiser, 2011; McKnight, 2006; Wheatman, 2001; Wheatman & Witty, 2001).

3 Methodology

The goal of this study is to understand whether digital repositories that have a preservation mandate are engaging in disaster planning activities, particularly in relation to their pursuit of trusted digital repository status. In cases where digital repositories are engaging with disaster planning, the study also examines the process of creating disaster response and recovery plans, with a focus on how these activities are integrated into the management of the digital repositories.

To answer these questions, the methodology of this study involves a mixed methods approach consisting of document analysis and semi-structured interviews to examine the disaster response and recovery planning practices of digital repositories.

This study was reviewed by the Institutional Review Board at the University of Michigan and was granted "Not Regulated" status.

3.1 Selection of Sites

The sample population for this study consists of digital repositories that have either sought trusted repository status, have conducted a TRAC, DRAMBORA, or DSA self-audit (and made the results of this audit publicly available), or have expressed a commitment to pursuing this type of certification process in the future. For the purposes of this study, trusted repository certification refers to the Trustworthy Repositories Audit & Certification (TRAC) as administered by CRL, or Data Seal of Approval (DSA) certification as the results of the certification audits for TRAC and DSA are publicly available, and the outcome of a successful audit is an official certification. The initial list of repositories was created in May of 2011, based on information available via their own websites, the Center for Research Libraries website, and/or the Data Seal of Approval website at that time.

As a result of the analysis of potential sites, an initial list of 19 organizations was compiled. In the end eight were selected for inclusion in the final study based on their availability at the time of the study and the willingness of individuals at those organizations to participate in the interview portion of this study. This initial list of 19 repositories was narrowed to the final group of eight based on the availability of respondents to participate in the one-hour interview portion of the study. All who were able to complete an interview by the end of January 2012 were included in the study.

The list of organizations considered for the study appears as Table 1, organizations that were included in the final group of eight are identified with italics:

Table 1: Initial List of Repositories

	Repository	URL
1	Archaeology Data Service	http://archaeologydataservice.ac.uk/
2	Archives New Zealand	http://archives.govt.nz/
3	*Chronopolis (The University of California at San Diego)*	https://chronopolis.sdsc.edu/
4	DSpace (at the Massachusetts Institute of Technology)	http://dspace.mit.edu/
5	ECommons (Cornell)	http://ecommons.cornell.edu/
6	*HathiTrust*	http://www.hathitrust.org/
7	*The Inter-University Consortium for Political and Social Research*	http://www.icpsr.umich.edu/
8	Library and Archives Canada	http://www.collectionscanada.gc.ca/
9	The Library of Congress	http://www.loc.gov/index.html
10	*MATRIX (Michigan State University)*	http://www2.matrix.msu.edu/
11	The National Archives and Records Administration	http://www.archives.gov/
12	The National Archives of Australia	http://www.naa.gov.au/
13	National Library of Australia	http://www.nla.gov.au/
14	*Portico*	http://www.portico.org/digital-preservation/
15	Statistics New Zealand	http://www.stats.govt.nz/
16	The California Digital Library	http://www.cdlib.org/
17	*The Internet Archive*	http://archive.org/index.php
18	*The MetaArchive Cooperative*	http://www.metaarchive.org/
19	The UK Data Archive	http://www.data-archive.ac.uk/

3.2 Discussion of Eight Sites

3.2.1 Chronopolis

Chronopolis is a geographically distributed preservation network that uses iRODS to "federate the partner sites and to replicate data among them" (San Diego Supercomputer Center [SDSC], 2011b, para. 3). "Originally funded by the Library of Congress, the Chronopolis digital preservation network has the capacity to preserve hundreds of terabytes of digital data—data of any type or size, with minimal requirements on the data provider. Chronopolis comprises several partner organizations that provide a wide range of services" (SDSC, 2011a, para. 1). The partner organizations that comprise Chronopolis are: San Diego Supercomputer Center (SDSC), UC San Diego Libraries, National Center for Atmospheric Research (NCAR), and the University of Maryland Institute for Advanced Computer Studies (UMIAC). "As of July, 2009, Chronopolis houses four diverse collections: a backup of the complete digital holdings of the Inter-university Consortium for Political and Social Research (ICPSR, based at the University of Michigan, 'Web-at-Risk' collections from the California Digital Library (CDL), geospatial data resources from the

North Carolina Geospatial Data Archiving Project, and several decades of data from research cruises from the Scripps Institution of Oceanography (SIO) at UC San Diego" (Minor, 2010, 121).

Chronopolis focuses on providing long-term preservation of digital resources. "Format obsolescence is not an immediate concern of the Chronopolis system. Instead, this is regarded as the responsibility of the data providers. The single, overriding commitment of the Chronopolis system is to preserve objects in such a way that they can be transmitted back to the original data providers in the exact form in which they were submitted" (SDSC, 2011c, para. 3). Users will be able to retrieve from Chronopolis exactly what they deposited with no changes to format or content.

Chronopolis received TRAC certification in 2012, with a final score of eleven out of a possible fifteen points (CRL, 2012a, 4).

3.2.2 HathiTrust

"HathiTrust is a partnership of major research institutions and libraries working to ensure that the cultural record is preserved and accessible long into the future. There are more than sixty partners in HathiTrust, and membership is open to institutions worldwide" (HathiTrust, 2012a, para. 1). The founding members of HathiTrust include the 12-univeristy consortium known as the Committee on Institutional Cooperation (CIC, #63) and the eleven university libraries of the University of California (UC) system (Rombouts & Princic, 2010).

HathiTrust is based at the University of Michigan's Ann Arbor campus, and is part of the University Library, with a mirror site located in Indianapolis, Indiana. The repository focuses both on preservation of and access to data. "HathiTrust Digital Library is a digital preservation repository and highly functional access platform. It provides long-term preservation and access services for public domain and in copyright content from a variety of sources, including Google, the Internet Archive, Microsoft, and in-house partner institution initiatives" (HathiTrust, 2012c, para. 1). The content of HathiTrust is primarily comprised of digitized monographs and serials from the participating member institutions.

HathiTrust has completed both TRAC and DRAMBORA audits. The repository received TRAC certification in 2011, with a final score of nine out of a possible fifteen points (CRL, 2011, 2). The results of HathiTrust's 2008 DRAMBORA audit are not publicly available.

3.2.3 Inter-University Consortium for Political and Social Research (ICPSR)

ICPSR is "an international consortium of about 700 academic institutions and research organizations" (ICPSR, 2011a, para. 1). The repository, which was founded in 1962, "provides leadership and training in data access, curation, and methods of analysis for the social science research community" (ICPSR, 2011a, para. 1).

Like HathiTrust, ICPSR is located at the University of Michigan in Ann Arbor. ICPSR, however, is part of the Institute for Social Research rather than the university library. Per the organization's timeline, ICPSR's first mainframe computer was purchased in 1967, the first

Digital Preservation Officer was hired in 2006, data backups were moved to spinning disk in 2008, and warm backup servers were deployed in remote locations in 2009 (ICPSR, 2011b). As an organization, ICPSR has a strong reputation for being a leader in the field of digital preservation and disaster planning. In addition to maintaining a set of publicly-available documents regarding digital preservation and disaster planning, ICPSR also administers training workshops in this area.

ICPSR participated in a test audit for TRAC, administered by CRL, in 2006. The results of this audit are publicly available via the CRL website (CRL, 2012b). ICPSR received Data Seal of Approval certification in 2010, the results of which are available via the DSA website (Data Seal of Approval [DSA], 2012; Data Seal of Approval Board, 2011).

3.2.4 MATRIX

MATRIX: The Center for Humane, Arts, Letters and Social Sciences Online at Michigan State University is a digital repository located at Michigan State University (MSU) in Lansing, Michigan. The repository was founded as part of H-Net in 1994 and today, "houses major digital library repositories including to the African Online Digital Library (AODL), Detroit Public Television's American Black Journal video archives, Historical Voices, and The Quilt Index. MATRIX also hosts the international scholarly networking community, H-Net" (MATRIX, 2012, para. 2). MATRIX is funded by a variety of sources, including MSU and various national grants and continues to maintain a focus on the humanities, arts, social sciences, and education.

Per the organization's website, MATRIX was, at the time of repository selection, working on writing several digital preservation policy documents, including a digital preservation policy framework, a digital preservation plan, and a disaster planning for digital assets document. As of March 2012 the website continues to reflect this intent.

3.2.5 National Library of Australia

The National Library of Australia "defines digital preservation as the processes involved in maintaining the required level of accessibility of digital objects over time" (National Library of Australia [NLA], 2012a, para. 2). A large part of the digital preservation effort at the National Library of Australia is PANDORA, Australia's web archive.

The National Library of Australia's digital preservation website includes a discussion of critical elements such as contingency planning and emergency response preparedness (NLA, 2012a). A version of the library's digital preservation policy is available online as well, including a statement that the library "stores and manages our digital collections in ways that will ensure their integrity, including adequate and secure backup and disaster recovery safeguards" (NLA, 2012b, sec. 6). In a policy statement covering the period of 2008 to 2012, the library's digital preservation policy states that by 2012 the goal is to be "well placed to prevent or respond to threats to the digital collections" (NLA, 2012c, sec. 2.1).

3.2.6 Portico

"Portico is among the largest community-supported digital archives in the world" (Portico, 2012a, para. 1). The repository works with academic institutions, nonprofit organizations, and

for-profit organizations such as publishers. As of March 2012 the organization has 739 participating libraries and 142 participating publishers Portico is a service of the nonprofit organization ITHAKA.

While Portico does not provide any specific disaster planning documentation through its website, there are several documents available through the Preservation Policies section that could be considered elements of a disaster plan such as, a succession plan, replication and backup policy, and escalation path for problem resolution.

Overall, Portico appears to be quite customer- and profit-focused in comparison to many of the other organizations in the study. The repository is not a part of any particular academic institution or national library, and the website provides information such as "How Portico Saves You Time and Money" (Portico, 2012b).

The repository received TRAC certification in 2011, with a final score of eleven out of a possible fifteen points (CRL, 2010, 2).

3.2.7 The Internet Archive

Founded by Brewster Kahle, the Internet Archive is a nonprofit organization that was founded with the mission of archiving and preserving the internet. The collection includes webpages, text, audio, video, and software, although the organization is perhaps best known by the general public for the Wayback Machine. The Internet Archive is based in San Francisco, California with a backup site that is located at an 'undisclosed location' that is also on the west coast (Internet Archive, 2012).

While the Archive is not affiliated with any particular academic institution, it appears to be less reliant on business from customer or member organizations than the other non-academic repositories included in this study. This is likely a result of the continued leadership and support of Brewster Kahle as the founder of the organization (Internet Archive, 2012).

In 2006, the Internet Archive's Archive-It program underwent a pilot assessment, which was administered by CRL. The result of the audit can be found via the CRL website (CRL, 2012b).

3.2.8 The MetaArchive Cooperative

The MetaArchive Cooperative was founded in 2006 as a membership of six academic libraries. The LOCKSS network has since expanded to include "libraries, archives, and other digital memory organizations" (Educopia Institute, 2012, para. 1). The Cooperative promotes a philosophy of encouraging institutions to preserve their own data rather than outsourcing preservation services to external vendors. This is accomplished by having each institution in the Cooperative maintain a server that is connected to the network (Educopia Institute, 2012). The geographically dispersed locations of the member institutions help to make the preservation more secure.

Documentation available via the MetaArchive Cooperative website a report of the results from a TRAC assessment that was carried out by an independent consultant. No score is available as this was not an official audit (Schultz, 2010).

3.3 Document Analysis

Documents for analysis were collected in two ways: Internet searching and asking participants during their interviews. The Internet, and specifically the websites of the repositories and their parent organizations, was searched for readily available documentation regarding disaster planning and digital preservation. At some organizations, such as ICPSR, this information was easy to find and to interpret. At others, such as Portico, it was much more difficult. Similarly, some interviewees were happy to share documents and information and others were reluctant, unwilling, or unable to do so.

Given the desire for openness and transparency among organizations who purport themselves to be trusted repositories, and organizations that have undergone TRAC certification and review, it was surprising that disaster planning documentation was so difficult to find, and in many cases not available at all.

Documents selected for analysis include any and all documents that have been identified by a particular repository as being relevant for their disaster response and recovery planning efforts. This includes disaster response and recovery plans, contingency planning documents, business continuity planning documents, succession planning documents, preservation planning documents, and TRAC and DSA audit reports, as well as documents for internal use only such as training documents and memos.

The availability of this documentation, and the types of documentation selected for consideration for this purposes of this study, help to provide insight to how organizations are engaging in disaster planning. The availability of this documentation also helps to show what information repositories make available in relation to their pursuit or demonstration of trusted repository status. Which is to say that some repositories seem to feel that making disaster planning information publicly available helps to cultivate trust from the community and others seem to place less importance on making this information available.

A complete listing of the documents included appears as Table 2 below:

Table 2: Available Disaster Planning Documentation

Repository	Documents
Chronopolis	TRAC Certification Report Digital Preservation Program Webpage
HathiTrust	TRAC Certification Report “HathiTrust is a Solution” Report

	"Building A Future By Preserving Our Past: The Preservation Infrastructure of HathiTrust Digital Library"
	Internal Planning Documents, Business Impact Analysis
ICPSR	Disaster Plan – Records/Finance
	Disaster Plan – Records HR (June 2007)
	Disaster Plan – Member Services
	Disaster Planning Resources (2008)
	Disaster Planning ICPSR Update (2007)
	Crisis Communications Plan
	Disaster Planning Policy Framework: Model Document
	Disaster Planning Roles and Responsibilities: Model Document
	Version 2.0 Disaster Planning Training: Model Document
	Disaster Planning Short-Term Action Plan
	Disaster Planning: Crisis Communications Plan
	Disaster Planning: Web Services Continuity Plan
	Data Seal of Approval Assessment Report
	CRL TRAC Audit Report
MATRIX	Information Security for Digital Assets at MATRIX
National Library of Australia	Building Trust: Pilot Preservation Audit of National Library of Australia Digital Repository
	National Library of Australia Request for Tender: Digital Library Infrastructure Replacement (RFT11103)
	National Library of Australia Collection Disaster Plan
	National Library of Australia, Digital Preservation Policy, 3rd Edition
Portico	CRL Report on Portico Audit Findings
	Portico TRAC Self-Report
The Internet Archive	Storage and Preservation webpage
	Petabox Storage System Information webpage
	CRL Archive-It Report
The MetaArchive Cooperative	MetaArchive Cooperative Charter

	MetaArchive Technical Specifications MetaArchive TRAC Audit Checklist (self-audit)

3.4 Interviews

Interview subjects were identified at each of the initial 20 organizations and were selected for inclusion in this study based on information available on the repositories' websites indicating that they are responsible for, or involved in, disaster response and recovery planning activities or digital preservation activities. In some cases multiple individuals were identified for a single organization. Interviews were conducted with ten individuals from eight different organizations between October 2011 and January of 2012.

A listing of participants interviewed is included in Table 3 below:

Table 3: Participants Interviewed

Code	Repository	Title/Role	Function
Subject A	Chronopolis	Digital Preservation Librarian/Project Manager	Digital Preservation
Subject B	Chronopolis	Project Manager	Digital Preservation
Subject C	HathiTrust	Assistant Librarian	IT
Subject D	HathiTrust	Digital Preservation Librarian	Digital Preservation
Subject E	ICPSR	Digital Preservation Officer	Digital Preservation
Subject F	MATRIX	Chief Technology Officer	IT
Subject G	National Library of Australia	Manager of Digital Preservation	Digital Preservation
Subject H	Portico	Archive Service Product Manager	Administration
Subject I	The Internet Archive	Director, Archiving Services	Administration
Subject J	The MetaArchive Cooperative	MetaArchive Program Director	Administration

Of those interviewed for the study, three hold administrative roles, five hold digital preservation roles, and two hold positions in information technology (IT). As the analysis will show, these roles are significant in that the subjects hold varying amounts of responsibility and authority within their organizations. Each also plays a different role in disaster planning activities within his or her respective organization. One limitation of this particular subject group is that only in two cases were multiple people at one repository interviewed. Specifically, Chronopolis and HathiTrust. A study that interviewed multiple people from different

departments and functions within each organization might be able to provide a more complete view of disaster planning at those organizations. For example, individuals in IT were able to speak about the technical side of preservation and administrators were able to speak about policy, but neither was well-versed in both.

Subjects were contacted via email, with a second follow-up message sent to those who did not respond to the first email. A total of 21 responses were received. Eight responders provided a referral to another individual within the organization, thirteen agreed to participate, and one declined to be interviewed but offered to answer questions via email. Of the twelve who agreed to participate, one was unable to schedule an interview in the timeframe allotted, another was willing to be interviewed but declined to be recorded or to share any documents, and the third had moved into a new position as was no longer involved in disaster planning. This produced a list of ten finalists who were selected to participate in the interview phase of the study.

Interviews were scheduled with those who responded expressing interest. Participants were sent consent forms to review and sign prior to the interview (see Appendix A). Interviews lasted approximately one hour each and were conducted via telephone or in person, depending on the location and availability of the subject. All interviews were recorded using two separate devices in order to ensure a reliable capture of the event. Post-interview notes were taken as well.

The interviews followed a semi-structured list of questions, which allowed each participant to discuss their own policies and practices, elaborating when appropriate. A semi-structured interview allows the interviewer to follow a predetermined list of questions, but allows for modification to the wording and/or order of those questions. The interviewer also has the ability to further probe particular areas in order to elaborate or clarify the subject's response (e.g. Babbie, 2010; Robson, 1993; Wildemuth, 2009). Questions asked covered the areas of: organizational attitudes toward disaster response and recovery planning, development of disaster planning documentation, access to disaster planning documentation, use and maintenance of disaster planning documentation, and budgetary considerations. Questions were based on a review of the literature and a preliminary review of the websites and available documentation at each of the 20 initial organizations. These questions fall into three areas: creation/development, access, and use of disaster planning documentation. For a complete listing of questions, please see Appendix B.

3.5 Interview Analysis

Once completed, the interviews were transcribed and coded using NVivo. The system for coding (or 'nodes') was developed based on a review of the literature, a preliminary review of the websites and available documentation, and initial impressions from the interviews themselves (e.g. Holsti, 1969; Wildemuth, 2009). Nodes fall into general categories of communication, documentation, administration, and preservation and are described in greater detail in Table 4 below.

Table 4: Description of Document Coding Scheme

Node	Description
Access	Access to the disaster planning documentation for both members of the organization and the general public.
Backup	Any mention of data backup used for digital preservation or disaster planning.
Backup Sites	Used only when a participant discussed backup via replication of data at more than one location.
Backup Tape	Used only when a participant discussed backup tapes (usually stored at a separate location).
LOCKSS	Used only when an organization is a member of a LOCKSS system/network.
Barriers or Difficulties	Barriers or difficulties to disaster planning activities. Usually barriers to creation or implementation of disaster planning documentation.
Budget	Discussion of how disaster planning activities are financed, how they fit into the budget. Also includes discussion of how disaster response fits in the budget.
Certifications	Discussion of audit and certification processes for trusted repositories such as TRAC or Data Seal of Approval.
Collaboration	Collaboration with external organizations in order to further disaster planning efforts.
Data	Discussion of the actual data held within the repository.
Data Loss	Discussion of specific events involving data loss, or of conditions under which data loss occurs.
Data Recovery	Discussion of specific events involving data recovery, or of conditions under which data may be recovered.
Disaster Events	Discussion of specific disaster events, real or potential.
Disaster Plan Documentation	Discussion of creation, development, and implementation of disaster planning documentation.
Internal Communication	Communication within the organization with regard to disaster planning, response and recovery activities.
Preservation	Discussion of digital preservation, not necessarily with a specific focus on disaster planning.

While this coding scheme was largely successful in highlighting the topics of disaster planning and audit and certification processes for trusted digital repositories, this method of document coding and analysis does have some limitations. For example, having only one researcher coding likely makes the results more subjective than they would be if multiple researchers were independently coding and comparing results. In addition, having the same researcher who conducted the interviews also carry out the coding introduces another layer of potential bias.

4. Findings

The combination of coded interview data and document analysis yield findings in six areas:

1. Incentive for Creation
2. Documentation
3. Process of Creation
4. Obstacles
5. Testing the Plans
6. Access to Disaster Plan Documentation

These categories have been influenced by the structure of the interviews, informed by patterns of analysis from NVivo, and are organized in an order roughly reflecting the chronological process of disaster plan creation and implementation.

4.1 Incentive for Creation

Many of the subjects expressed the idea that the development of disaster planning policies and procedures happened as a result of growth and development of the repository. However, even those who insisted that their repository had disaster planning policies and procedures in place before going through an audit certification process indicated that it was only through the process of responding to the needs of the auditors that they actually created their formal disaster response and recovery planning documents.

Subject B from Chronopolis stated, “everybody kind of knew, but you had to be part of the Chronopolis team to know what the disaster plan was. Now during the TRAC audit . . . Chronopolis itself has had to make that more public.” Subject A from Chronopolis echoed this statement, “it [the TRAC audit] really did push us to create a lot of documentation and to be very explicit about things that we had just kind of assumed before or that we hadn’t put into place or had language for.” And in addition to both acknowledging that the TRAC audit provided the incentive to formalize information that had been informally or tacitly understood within the organization, they each also stated more explicitly that “the main reason that we wrote specific documents was as part of an audit process which began about the middle of last year” and that “we only documented all of this because of TRAC.”

The MetaArchive Cooperative is an organization that has gone through an internal TRAC audit. Much the same as Subjects A and B from Chronopolis, Subject J stated that disaster planning policies and procedures had been in place prior to the audit, “I would say in some ways the disaster planning action has been in place since 2004, since we first brought up the network.” He also stated that, “there is a second set of documentation that we prepared in response to a TRAC audit that we did in 2008,” adding that, “it's [TRAC] very good at crystallizing and condensing down what things you should be documenting and it gives you a good base in my experience for defining and making sure that your practices are as sophisticated as they need to be in order to guarantee that you're doing digital preservation . . . the disaster recovery piece is a perfect example because that document and the succession planning document those have come out of that TRAC experience, not because we hadn't already thought through those things

and had them documented in other ways, we did not have one document that said 'this focuses completely on that topic' and that, the importance of that, was highlighted in the TRAC document and I think rightfully so. It helped to motivate us." Again in this case, it was the audit process that provided the incentive to create the documents that are currently considered to be the organization's disaster planning documentation.

Subject H from Portico stated that, "growing size was the big impetus . . . I don't think that there was a situation that caused us to think 'oh my goodness we need a set of plans.' It was more, 'wow we're getting really big and if we ever have to recover we want written down the steps we're going to do,' so it was just environment more than anything else . . . As the organization grows, and the content grows, these things become more urgent in part for the very practical reason that the reality is that if a disaster were to occur as the organization has grown and gotten bigger, the trauma of recovering gets bigger and so the very practical needs to have a plan in hand become more urgent." Even more than the previously mentioned repositories, this interview emphasized the fact that the disaster response and recovery planning policies and procedures were created out of a recognized need to support the growth of the repository. However, even in this case, the interviewee went on to explain that, "our final policies were drafted in such a way to try and answer specific questions in TRAC . . . we didn't really write anything specifically to meet TRACs needs, but . . . when we sat down to write policies it was really just writing policies that matched the decisions we were making . . . for the purposes of CRL and their TRAC-type audit, we spent quite a bit of effort framing the policies and all of our other documentation in such a way as to help it answer TRAC questions . . . it was an ordeal, it was not easy." The staff of this repository, like Chronopolis and The MetaArchive Cooperative, ultimately created disaster planning documentation based on what was needed for their audit.

While the interviewees from HathiTrust acknowledged during the interview that the repository does not have formalized disaster response and recovery documentation in place yet, Subject D stated that, "a lot of where we are in the planning is documenting what we're already doing, creating formal policies that describe what we're doing, and this was a large part of our TRAC certification also." In this case, their position seems also to be that the policies and procedures exist and the process of documenting these policies and procedures is something that is being undertaken specifically as a result of the audit.

Of the interviewees from organizations that have undergone an audit for certification, Subject E from ICPSR is the only one that did not report the development of specific disaster planning documentation for the audit. Much like the others, Subject E developed disaster response and recovery policies and procedures as a result of organizational growth and development, "I think it was just a general sense of alignment with good practice. The reason that we started talking about it at all was that there was a sense that we need some kind of disaster planning in place." Additionally, "we had three near-miss situations or things that would be considered at least emergency situations, none of which we could have described ahead of time." However, the result of this decision to develop these policies and procedures was the creation of formalized documentation, including templates for use by other organizations, that the staff of ICPSR has

made available via the Disaster Planning section of their website. Rather than following a format based on the needs of an audit, they relied on "the NIST model."

Staff from the remaining repositories (MATRIX, the Internet Archive, and the National Library of Australia), all discussed the existence of disaster planning policies and procedures, generally as documents or checklists that reside within the IT department. It seems that these repositories have met the first step as described by the other organizations discussed in this section, but have not yet taken the step of formalizing these policies and procedures as a set of dedicated disaster planning documents. One could speculate that if these repositories went through an audit they would likely complete the process by creating these documents.

The theme of formalized disaster planning activities for the purposes of audit or certification was prevalent throughout the discussions with those repositories that have been through some form of audit. Accordingly, the documentation coding scheme was able to capture the overlap between respondents discussing certification and disaster planning documentation. As can be seen in the chart below, institutions that have undergone certification audit have a higher level of overlap between discussion of certification and disaster planning documentation as well as a higher instance of discussing each independently.

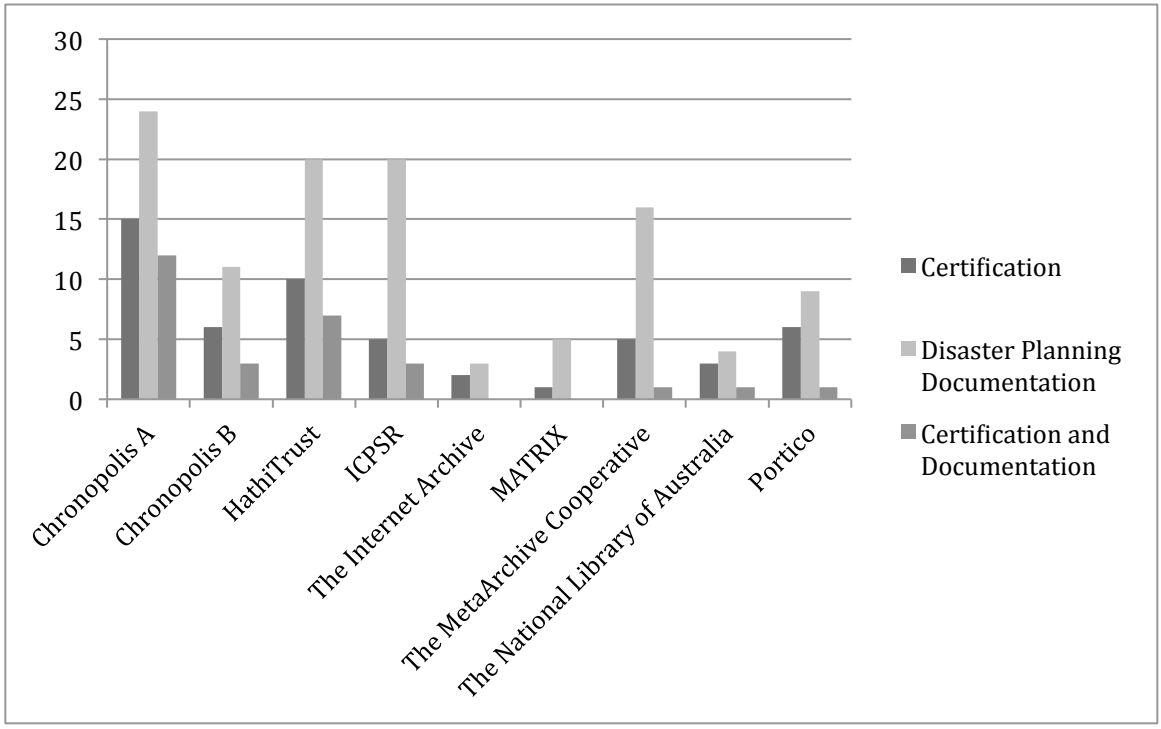

Figure 1: Certification and Disaster Planning Documentation

While TRAC is only one of several possible audits that the repositories in this study have undertaken, TRAC certification is the most commonly held type of certification. Among my respondents, three repositories are TRAC certified, one repository is DSA certified, and several others have completed informal self-audits. Of the three repositories that have been TRAC

certified, the two with the highest scores report the most complete disaster planning documentation.

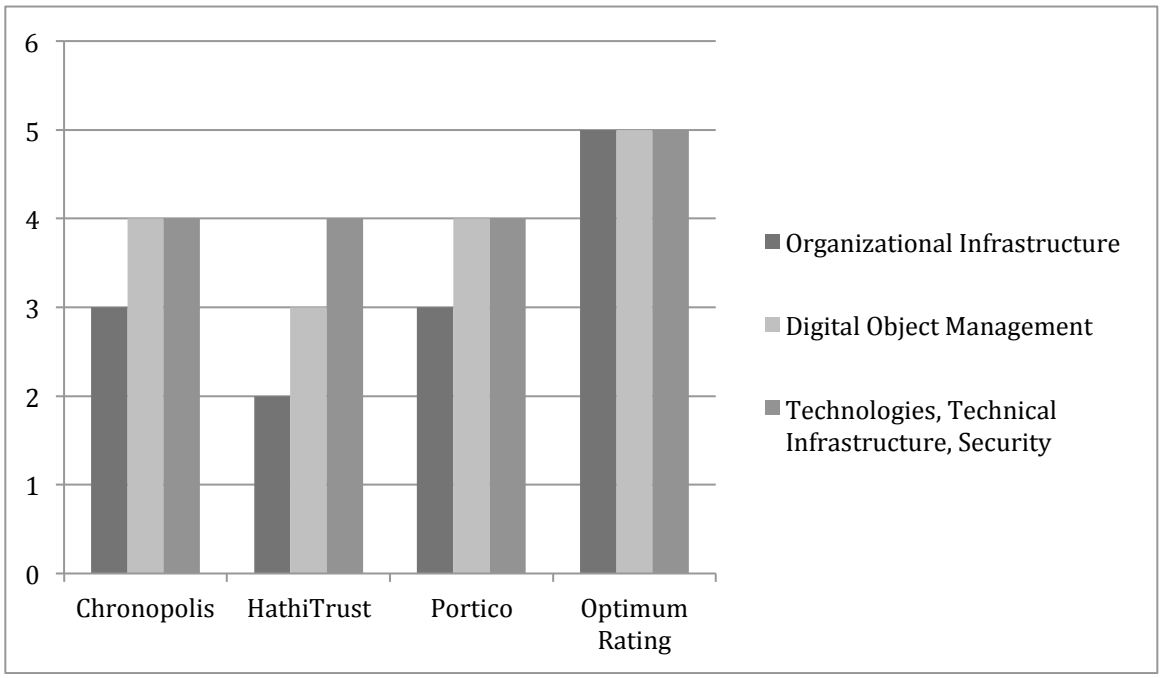

Figure 2: TRAC Audit Results

4.2 Documentation

While all of the interviewees were quick to provide assurance that their organization did indeed have disaster planning documentation, and many were happy to provide evidence of that documentation in the form of an audit report, very few were able or willing to discuss these plans in detail or to provide copies of the complete documentation.

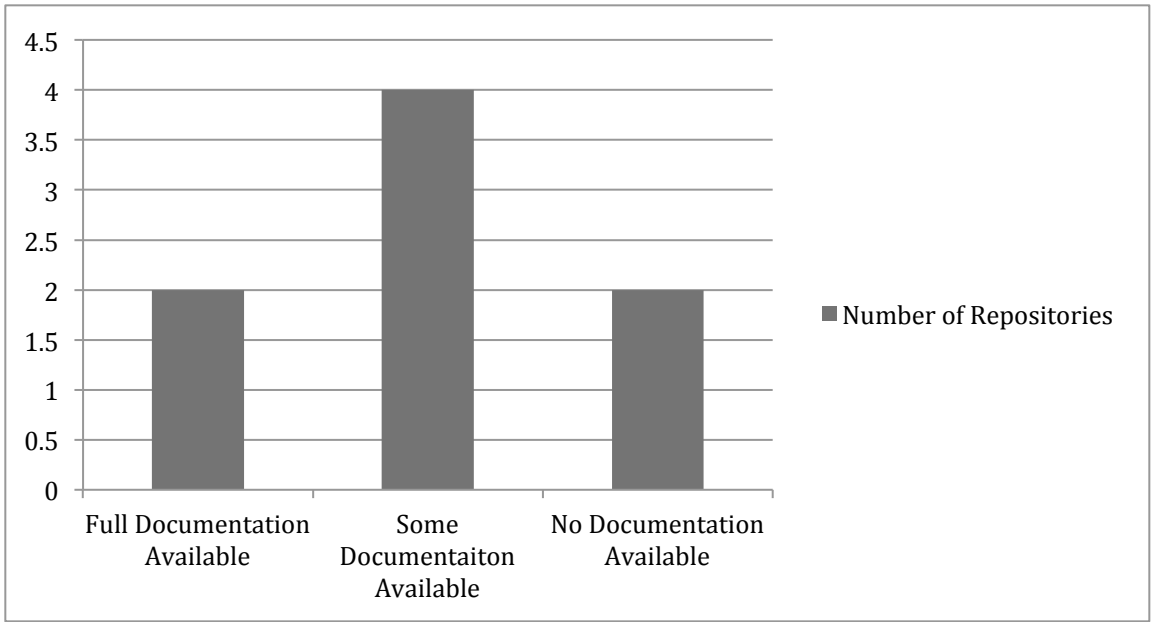

Figure 3: Disaster Planning Documentation

Both interviewees from Chronopolis explained that the organization's disaster planning documentation was created in response to the recent TRAC audit. Per Subject A, "in general terms we created a TRAC report which basically follows the question and answer schema of the TRAC audit itself." But upon further discussion, Subject B revealed that the Chronopolis disaster plan primarily serves to point users to other disaster planning documents, "we do have a document that is Chronopolis disaster planning, but all the instructions for that disaster planning link out to other places." Specifically that, "Chronopolis is a consortium of three institutions . . . and each of those entities has a specific disaster plan for what happens to data in their data centers. And so we rely on those disaster plans in those data centers to make up the whole disaster plan for Chronopolis." Subject B discovered while reviewing documentation during the interview that, "it's just a statement, we don't actually link out to the other institutions." Meaning, the information that is available to the public references other documents but does not provide links to those documents. In the words of Subject B, "they're actually difficult to find. So yes, they're available to the public - but they're available if you can find them." The interviewee explained that he was actually unable to find the documents without assistance from the individual responsible for them at the parent organization.

Subject F from MATRIX stated that, "we have practices and we have some documentation in different locations that more or less equate to that [disaster planning] but we don't have a direct formal plan that speaks to exactly what we'll do in the event of a disaster." In an interesting exchange, Subject F explained that a formal disaster plan is not needed because the steps required to recover from a disaster event are so obvious and simple that any competent System Administrator would understand how to carry out this action. "We have a wiki and we've been putting a lot of our documentation on that. And we do have a lot of our documents regarding how to bring the system back up, and what our plans are, and what our procedures are. They're not in one actual spot on the wiki yet but we're getting to that point, and really

part of the decision to make with us is do we focus on documenting more or less a known procedure . . . most Sys Admins would understand 'ok there's a tape backup, take the tape backup and restore it' and now you're good to go more or less. I mean if at worst case someone hopefully would know to put a tape in the drive, right? It's common sense . . . at that point it's really a question of to what level of detail do we get . . . but documentation we haven't really focused on a lot just because of the fact that we're not at a point where we're complex enough to require it in my opinion." This is an opinion that was not expressed by any other subject in this study, and which may be a result of the fact that this subject was in a role with an IT function rather than a role with a preservation function. As will be discussed later, many interviewees discussed having difficulty in getting proper documentation from the IT departments within their respective organizations. This discussion perhaps provides some insight to the other side of that frustration.

Subject J from The MetaArchive Cooperative explained that, "we have documented contingency plans that look at a number of different points on the axis of problems that could erupt and what would happen in those kinds of disaster scenarios" and also that, "there is a contingency piece for each one of our member institutions that is part of their own disaster planning so there are these two layers to disaster planning as we see it at MetaArchive." In this case, there are multiple components to the disaster documentation. Subject J goes on to explain that "in terms of documentation, it started with our membership agreement and our charter and those two core documents are the legal underpinnings for the relationships that comprise the MetaArchive network" as well as "a second set of documentation that we prepared in response to a TRAC audit that we did in 2008 that [resulted in] a formalized contingency plan document and succession plan." In this case, the interviewee was able to discuss and describe several types of documents, but again specific disaster planning policies and procedures were not available.

The staff of Portico have created several different documents that comprise their disaster planning documentation. Per Subject H, "our policies are very targeted, so we don't have one big overarching policy for Portico. We have a series of smaller policies . . . so at Portico we've got 13, 16, 21 different policies . . . I would say that there are probably three policies that are directly impacting disaster recovery." This interview also revealed that, "we have two sets [of disaster planning documents]. There is the set that is maintained by our IT group . . . they have a set of disaster recover policies that they have developed that involves a lot of this infrastructure type stuff. Portico proper has a set of group preservation policies around disaster recovery, which specify the number of backups we need to have the number of replicas where they're going to be located, our general philosophy about it." Much like the interviewee from MATRIX, Subject H from Portico seems to be describing a disconnect between the IT and Preservation functions within the organization. Rather than having one set of combined disaster planning documents, there are two separate sets of documents that do not seem to be combined in any formal or significant way. In fact, the interviewee was able to discuss the IT documents in only broad strokes.

The staff of HathiTrust were totally open in terms of sharing their documentation and work in progress. As Subject C stated, "it's in progress right now. We have a foundational outline that we're working from" and "it's not a functional recovery plan by any means but the goal is to get to that." This discussion reinforced the idea that the policies and procedures needed for actual disaster response and recovery are in place, and that the creation of formal disaster planning documentation is a formality, "a lot of the proper thinking has been done in very many ways and the proper work has been done to ensure that things will likely function very smoothly in the event of a disaster, but the work has not been done to fully articulate the processes in which it will take place."

The staff of ICPSR were also completely open about sharing their disaster documentation. Subject E stated, "we follow the NIST model for the types of documents. So it's a suite of documents it's not a single thing. It's ongoing, it's a planning process, the focus is on planning as a verb, not plan as a noun." Nearly all of their documentation was available via the disaster planning section on the organization's website, and the individual interviewed was able to share the remaining documents via email.

The staff of the Internet Archive have, "an internal checklist absolutely which we review" that is maintained by an IT department. However, as discussed above, this checklist is not considered to be a complete disaster response and recovery plan. It is also not available to the public in any form. Subject I in this case was either unwilling or unable to discuss specifics of this plan. Subject G from the National Library of Australia expressed a similar situation, "we have a digital preservation section, and we have a very large IT section, and the IT section deals with a lot of the things like backups . . . and so as much as I could say to you 'yes we do have a backup regime' I can't give you the exact details of it because they run those kind of things." The disconnect between the preservation and IT functions at this organization is so great that the digital preservation section is actually not familiar with the disaster planning documentation at all.

4.3 Process of Creation

One of the principal findings of the study is that the process of seeking certification is extraordinarily time consuming and requires a major commitment to documentation. The creation of disaster planning documentation is reported by most repositories to have been one of the most time consuming aspects of this process. Discussions with subjects focused much more on the investment of time and people into the process of creating disaster planning documentation than on the specifics of decisions made within those documents.

For Subject A at Chronopolis, "it was the audit that was 100% our guiding force in creating these documents." And specifically, they "took the requirements that they [TRAC] had listed, put them in a big excel spreadsheet and then used that to drill down into specific questions." Subject A was also able to discuss some of the individuals involved in the process, although not in great detail, "internally we had a couple different individuals contribute, the primary one obviously would have been the Data Center Manager, who was in charge of maintaining the

equipment." He was also able to speak to the creation of disaster planning documentation at the three partner institutions, "and similarly that was the case at all three institutions that those typically were the people involved in creating the documents needed for the audit." The documentation took "a good three to four months" to complete, and "was one of the more significant sections that we had to do a lot of new work for . . . it's probably one of the larger sections for us in terms of how much time was spent on it."

The discussion with Subjects C and D from HathiTrust focused on both how they are proceeding, and on how they expect to create their disaster planning documentation. The general philosophy of the organization expressed by Subject D to "do very much what's practical and try not to predict the future" and the interview with subjects from HathiTrust reinforced this idea. Specifically, Subject D stated that, "we take our practices and are not constrained by a policy necessarily until we actually do that thing." Meaning that the goal of their disaster planning document creation process is to articulate current practices rather than setting rigid policies and procedures that will need to be implemented and enforced.

Subject J from the MetaArchive Cooperative focused on the amount of time spent completing the disaster planning documentation. "For the initial investment, when you look at all of the different people who are involved and all the different stages of that drafting, I would say at least 80 hours of people time went into the drafting. Not the approval process, not the continued revisions that we're still doing, but just the base-level drafting to really get all of this done and lined up . . . at least 80 hours." Subject J, however, did not discuss the specifics of how that time was spent, or of whose time was spent.

Subject H from Portico also talked about length of time to complete the disaster planning documentation, although it was discussed only as part of a larger project to document digital preservation policies. "I don't know specifically around disaster recovery, I would say that it was probably a six month process for us . . . we probably took about six months to really formalize and finalize a relatively substantial set of our preservation policies, disaster recovery being one element of that the whole process . . . it was actually quite a chunk of time with participation from three or four people. It was not an easy process." Similar to HathiTrust, the staff of Portico based their disaster planning documentation decisions on factors outside of the TRAC audit process, focusing on three specific elements, "environmental review of what is recommended, some very practical considerations about what is physically possible, and then some business considerations about marketing and outreach." Also like HathiTrust, the staff of Portico expressed the organizational attitude of crafting policies and documents that meet their needs and describe their practices without being overly prescriptive or restrictive, "we have a standard template for how we write policy documents. Our policy documents tend to be . . . relatively high level and strategic, which allows us to write and change implementations over time and have different implementation documentation to support the high level strategy." In this case, Subject H is again confirming the previously discussed finding that repositories are reluctant to discuss disaster preparedness documentation in great detail.

4.4 Obstacles

The majority of interviewees included in this study reported significant obstacles or challenges encountered in the process of creating their disaster response and recovery planning documentation. The most common themes were the difficulty of getting buy-in from other members of the organization, difficulty collaborating and communicating with the IT department, and the amount of time required for completion of the documentation. These obstacles align with the previous finding that most repositories that have formalized disaster planning documentation created that documentation as the result of an audit. In other words, they were unable or unwilling to create the documentation without an organizational mandate to do so.

The chart below indicates the number of times the topic of barriers or difficulties to disaster planning was mentioned in each interview. This chart shows that the interviewee from the organization with the most complete and publicly available disaster planning documentation (ICPSR) also spent the most time discussing obstacles to disaster planning. The repository with the next highest incidence of discussion of obstacles is HathiTrust, a repository whose staff has yet to complete their disaster planning documentation.

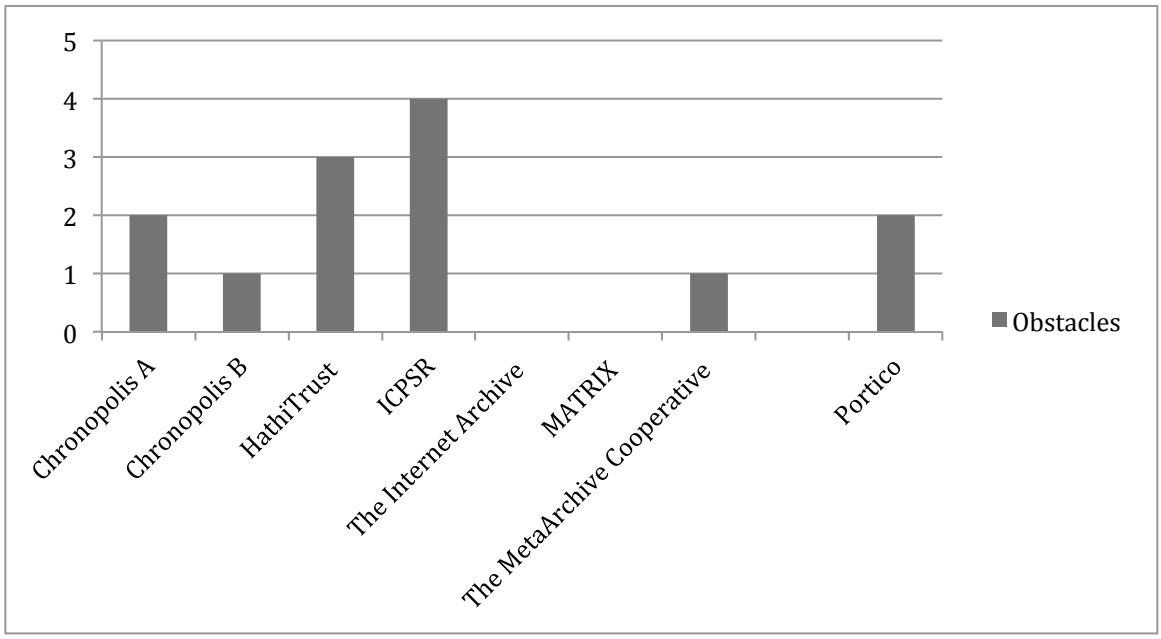

Figure 4: Obstacles

Subject A from Chronopolis described the process of creating the disaster planning documentation as "herding cats." He went on to say that the most significant barrier to completing the disaster planning documentation prior to the TRAC audit "probably would have been not having a big enough stick to force people to do it . . . it wasn't until we had an auditor come in that we said, 'ok look here guys that's not good enough or if that is what we're going to get, that's definitely going to reflect on our audit report.' And so I have to say for us, and I don't want to overstate this because we didn't have a problem before, but in order to get detailed documents it really did take the audit to pull those things out." Chronopolis staff were unable

to get detailed disaster planning documentation from the three partner organizations until the audit report imposed a higher degree of accountability.

The interviewee from ICPSR, the only organization in this study with staff who created detailed disaster planning documentation independent of an audit, focused on the problems of organizational cooperation and difficulty coordinating with the IT department. Subject E began with discussion of the historical resistance to formal disaster planning activities, "in the past I think that it was often looked at as a luxury . . . it's a natural human thing to not want to talk about a disaster until the disaster is there and then be caught short because you don't have any planning in place." In order to overcome this resistance, it was necessary to get buy-in from senior members of the organization, members who were initially unwilling to devote their time to the process, "part of the difficulty of engaging in roles and responsibilities is that they have to at least start at the highest levels of the organization. They view it as costly they view it as a distraction, but you can't work at the bottom when you're dealing with decision making and actual authority." Staff of ICPSR also experienced difficulty in "parsing out the IT piece . . . because when you have IT as an integral part of your organization and your organization is committed to lifecycle management, there is this 'now' and 'future' and the people who are doing these things don't often distinguish between the hats that they have. It was hard to get them to focus on the different parts . . . we have a really good IT group, but it's also a challenge for digital preservation."

Subject J from the MetaArchive Cooperative focused more on the problem of time, "it is all about time, it takes so much time to write documents, figuring out what goes in the documentation and getting it by the committee, and then starting the draft and getting that past the committee and getting it approved . . . I would say that time is the greatest challenge of everything that we're doing around policy creation, including disaster recovery planning." This also indirectly discusses the problem of organizational buy-in and cooperation as it is the process of getting committee approval that seems to take the most time.

Subject G from the National Library of Australia expressed the greatest degree of frustration in coordinating with the IT section of the organization, "we have certain people in our IT section [who] I'll say are preservation deniers." He then goes on to explain that, "sometimes our IT department 'know best' - they take the high moral ground and then we catch them out and say 'hang on a minute, you're not doing this, or this' . . . and then IT says 'we're working on it' and we found some frustration because they are a separate department." While formal disaster planning documentation is not currently in place at this organization, this difficulty in coordinating with the IT section in order to document and carry out digital preservation activities extends to disaster planning as well.

Interviewees from HathiTrust were the only to explicitly state that there were no problems getting organizational buy-in. Per Subject C, "there is very little, and maybe safe to say no, organizational resistance to . . . making the data as safe and as useful as possible . . . organizational philosophy is to make sure that all of the work that has gone into this data is not going to be for naught when it all gets wiped out in some sort of disaster." However, this is an

organization that has gone through an audit but has not yet produced formalized disaster planning documentation. As Subject C stated, they are struggling with the challenge of producing documentation that is both excellent and useful and are working to embrace the philosophy of not "let[ting] the perfect be the enemy of the good." This case seems to highlight the value of the obstacles that the other organizations have faced.

The figure below illustrates the discussion of obstacles, of documentation, and the overlap of obstacles and documentation. Interviewees from three of the four repositories that show an overlap in discussion of obstacles and documentation have successfully completed certification with either TRAC or Data Seal of Approval and report that they have formalized disaster planning documentation in place. The fourth, HathTrust, has also successfully completed TRAC certification, but does not yet have a formalized disaster plan.

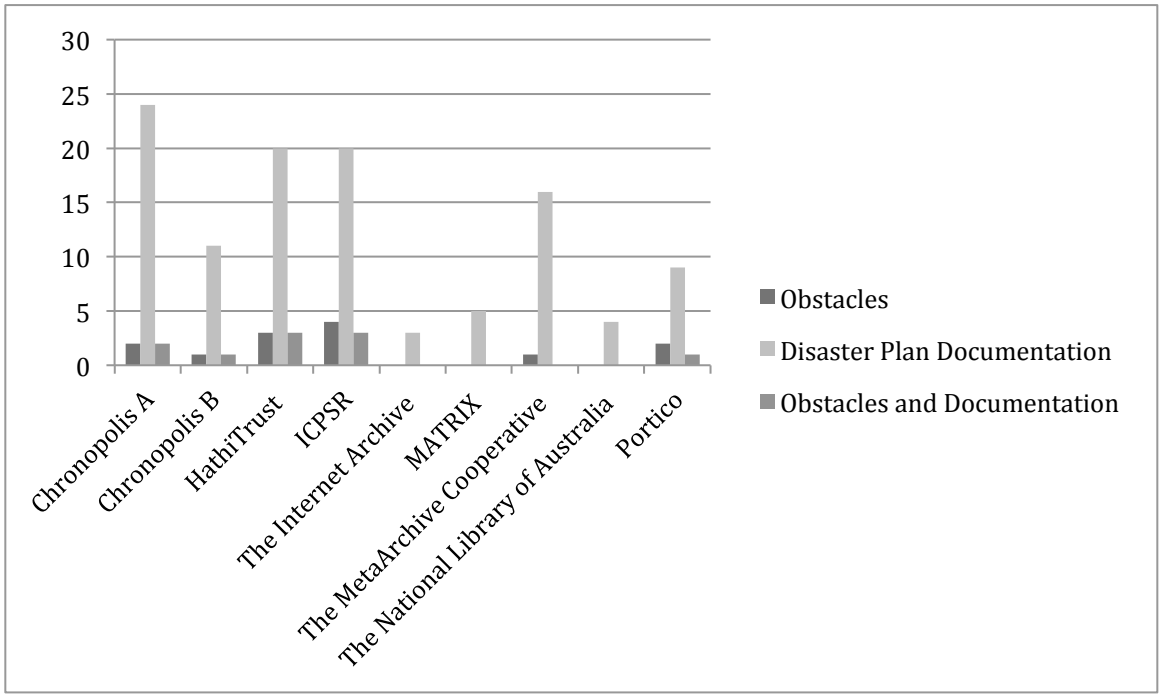

Figure 5: Obstacles and Documentation

4.5 Testing the Plans

Another theme that emerged from the interviews was that of testing the disaster planning documentation through exercises or "fire drills" once the repository had a formal disaster plan in place. Three of the eight repositories studied discussed some form of exercise to test the plans, and one explained that they chose not to run these drills because they encounter enough issues on a regular basis that their policies and procedures are under constant scrutiny.

Interviewees from Chronopolis, ICPSR, and The MetaArchive Cooperative each discussed their efforts to regularly test their disaster planning documentation. For Subject A at Chronopolis, that involves testing wherein "we basically go and unplug systems and then force ourselves to recover." Subject J at The MetaArchive Cooperative also discussed "disaster planning exercises" that are conducted "on a two year basis." However, the representative from The

MetaArchive Cooperative stated that, “we haven't run a full scale again . . . we’ve done pieces of it at different times.” Both of these interviewees explained that they first ran the testing exercises for the purposes of the TRAC audit. The staff of the MetaArchive Cooperative have not run another full drill since that time and Chronopolis has not yet had the opportunity to run another drill since their audit was conducted so recently. It will be interesting to see if they continue with a regular testing schedule.

Subject E from ICPSR described an elaborate tabletop exercise scenario in which, “somebody who is in web services in our organization [who] is really involved in gaming created this whole scenario for us and handed out cards of who are you and what's your part of the scenario.” From this initial exercise they were able to identify weaknesses in their plans, “somebody had to go away and fill in some gaps.” As a result, the organization has “committed to doing that at least once a year.”

Subject F at MATRIX, who is in an IT role within the organization, stated that “we have enough issues from time to time that we don't run fire drills anymore because it happens and a lot of it has to do with upgrading our infrastructure . . . it's just really when there's underlying core systems that get upgraded that's when the issues could occur.” In other words, they spend enough time fixing problems that they know how to recover. This response echoes other findings regarding the relationship between preservation and IT. Namely, that it is those individuals whose roles fall under preservation functions who drive the development of formalized disaster response and recovery planning documentation, and that individuals whose roles fall under IT functions are more resistant to these types of activities.

4.6 Access to Disaster Plan Documentation

Most subjects in the study expressed a desire to make their organization’s disaster planning documentation available to the public at some point in the future. To date, ICPSR and HathiTrust are the only repositories whose staff have made all of their documentation publicly available via their website or via email request. In terms of internal access, most repositories have disaster planning documentation that is available to all employees of the organization via some sort of staff wiki or shared storage space.

Subject E from ICPSR indicated that approximately 80% of their disaster planning documentation is available on their website, and that “we try to be really transparent and if people ask us really we would provide anything else that isn’t online.” Subject E did, in fact, share additional information via email for the purposes of this study. The staff of HathiTrust has placed the foundation documents for their disaster planning efforts online via the organization’s website, and one of the interviewees stated that they intend to continue making additional documentation available, including the completed disaster planning documents once they are completed. Their position on the matter, according to Subject D, is that “it’s only to our benefit I think to make as much of that available to people so they know what we’re doing . . . our orientation is open.” For the purposes of this report, they were able to share additional documentation and information via email.

Subject A from Chronopolis explained that, “we do intend to make it as public as possible,” but also stated, “we're going to have a couple points where whatever we make public we're going to have to do it at such a high level that it doesn't reveal compromising details . . . and not put anything there.” The organization is struggling with wanting to make their information available but needing to respect the privacy and security needs of their partner organizations. These are the same partner organizations that were reluctant to even provide disaster response and recovery information to Chronopolis to begin with, as discussed above. They did, however, indicate that the entire staff and all partner organizations have access to this information internally.

Staff at all member institutions forming the MetaArchive Cooperative have access to their disaster planning documentation for all staff. Per Subject J, “that's all on our wiki, which is part of our core infrastructure that is housed in the cloud so it is accessible to all of our member institutions both for maintenance of the document and then also for viewing the document. It is a password protected area.” Access to this documentation is restricted to “members or people who have formalized relationships” with the organization. The staff of the MetaArchive Cooperative is in the process of trying to make this documentation available via their website, and Subject J stated that “for the disaster recovery thus far it does not look like we'll need to restrict the information therein, there's nothing there that . . . would compromise MetaArchive.”

Portico, MATRIX, and the Internet Archive all have documentation that is available internally to staff but do not plan to make that information available to the public. Subject F stated that at MATRIX, “all students and staff, anyone that basically has a MATRIX login which is everyone that works here [has access to the documentation on the staff wiki].” Subject H from Portico maintains that, “the IT specific disaster recovery documents, which are pretty specific and not the type of thing you can make publicly available, are maintained internally only.” And while Subject I from the Internet Archive stated that, “everyone [in the organization] has access to it [the checklist]” and that “one of the founding principles of the Internet Archive is universal access to all knowledge . . . that trickles down to all of our documentation,” he also stated that he would not be able to share their disaster response and recovery checklist with anyone outside of the organization.

Subject G from the National Library of Australia was the only person who communicated that the disaster planning documentation was not available publicly, nor was it available internally. Rather, the documents reside with the IT department and have not been shared or made widely available, “I haven’t [had contact with the plan] . . . our new systems hopefully will make it available to everyone.” This echoes the sentiment expressed by Subject G earlier in the interview, and discussed elsewhere in this paper, that one of the primary obstacles to disaster planning in this organization is communication with the IT department.

As Figure 6 below shows, the interviewee from the repository with the highest level of access to documentation, ICPSR, also discussed access to documentation more frequently. Staff at the

repository with the lowest level of access to documentation, the National Library of Australia, mentioned access to documentation the least. Other organizations whose interviewees also report low levels of access to documentation such as MATRIX and the Internet Archive also show low rates of discussion of access to documentation. Interviewees from repositories that provide access to some, but not all, of their documentation generally fall somewhere in the middle.

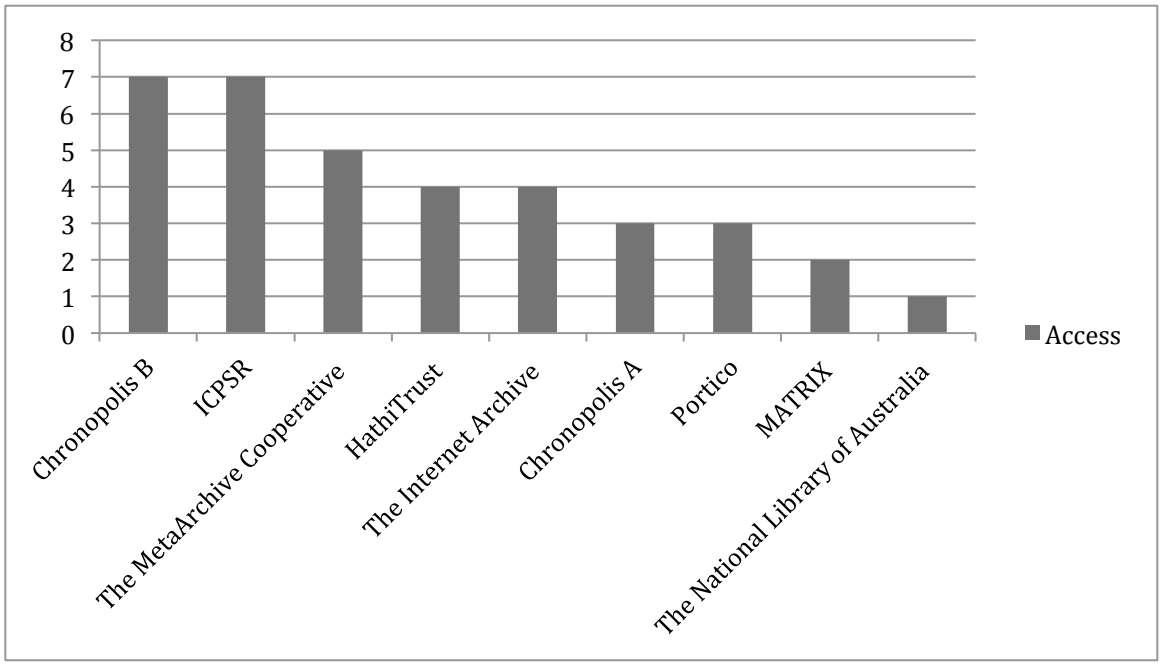

Figure 6: Access

5. Discussion

In this section, I will discuss the three main findings of this study in greater detail. The three major findings of this study are that:

1. For most organizations, the process of going through an audit for certification as a trusted repository provided the impetus for the creation of formalized disaster planning documentation.
2. Despite the desire for openness most repositories struggle with making their disaster planning documentation publicly available.
3. The single greatest obstacle to disaster planning activities at all stages of the process is coordination between the IT and preservation functions.

I will also address possible limitations of the study, and discuss directions for future research.

One pattern I observed throughout the interview process was that organizations such as Chronopolis, ICPSR, Portico, and The MetaArchive Cooperative, that have been through some sort of audit process, were more likely to have complete disaster planning documentation in place. These organizations discussed the role that the audit played in providing motivation to

complete this documentation, and discussed the challenges that had prevented them from completing this documentation previously. Central to this was the idea that until the organization was provided with a suitably attractive incentive (i.e. certification), it was difficult or impossible to convince other departments such as IT and Administration to spend time documenting policies and procedures that were either formally documented elsewhere or tacitly understood.

Repositories struggle with the decision to make disaster planning documentation available to the general public. I expected repositories that had been through an audit for certification of some sort would be willing to make at least parts of their disaster planning documentation publicly available. As one of the value principles of TRAC is transparency, and this transparence helps people to trust repositories, it seemed natural that they would then make that same information, or some portion thereof, available to the public. As I discovered, this is generally not the case. The availability of documentation regarding disaster planning activities varies widely among repositories and runs the full spectrum from fully available to completely restricted.

Finally, I found that the single greatest obstacle do disaster planning activities at all stages of the process is coordination, or lack of coordination, between the IT and preservation functions within an organization. Subjects in preservation and administration roles expressed frustration with the lack of communication and cooperation from the IT departments in their organizations. Subjects in IT functions expressed a belief that formal disaster planning activities were unnecessary and a poor use of time and resources for the organization. This is related to the first finding in that it seems to be the case that organizations are best able to overcome this obstacle are those that can demonstrate a concrete benefit, such a certification, that will result from the production of formal disaster planning documentation.

These findings suggest that one of the primary benefits achieving trusted digital repository status, in addition to the certification itself, is the fact that it provides an incentive for the entire organization to create accurate, up-to-date, thorough documentation of policies and procedures. For organizations that already have documentation in place, such as ICPSR and Portico, the audit provides the organization with an opportunity to improve and update their documentation.

These findings also suggest that a greater degree of communication and cooperation is needed between preservation and IT functions within digital repositories. A consistent pattern in the interviews was the difficulty in working with IT, and the resistance of that group to participate in formal disaster planning documentation efforts. Conversely, this problem can be seen as a shortcoming on the part of digital preservation policy makers. Perhaps an opportunity for education and better communication exists between the different functions. While the IT function seems to almost universally have been an obstacle to disaster planning efforts in the repositories in this study, interviewees also stated that this seems to be a case of individuals in the IT role not having the same understanding of and appreciation for disaster planning. An opportunity exists for those in the field of digital preservation to find ways of communicating

with those in IT, in order to improve collaboration and coordination throughout the organization.

The initial research question for this study focused on investigating how repositories are engaging in disaster planning activities. After examining the practices of several well-respected digital repositories, it has become clear that one of the reasons that so few studies have been conducted in this area is that digital repositories, until recently, did not to have documented their disaster planning efforts at all. It has also become clear that it is not possible to gain a full understanding of the disaster planning efforts of an organization if those efforts are not codified and made available for review. The fact that only two of the eight repositories were able or willing to make their disaster planning documentation publicly available was a major limitation for this study. Additionally, this lack of models may be hampering disaster planning efforts in the community

There are several other factors that could be considered weaknesses or limitations of this study. First, the small population size makes it difficult to draw conclusions that could be generalized to a larger population. This limited scope is partly a result of the small number of repositories that are engaging in trusted repository audits and partly a result of the limited timeframe in which this study was conducted. Additional studies in this area may want to consider including a greater number of repositories.

Second, speaking to only one or two individuals at each repository does not provide a complete picture of the entire lifecycle of the disaster planning process. In order to gain a full, complete understanding of the activities at any given repository, it would be ideal to interview several individuals from different departments or functions within a repository. For a study such as this, interviewing individuals in digital preservation, administration, and IT would provide a well-rounded view of disaster planning activities.

Third, future research would do well to either focus on one type of repository, or to study a broad spectrum of repository types. Of the final eight repositories included in this study, one is a national library, one is an institutional repository, and the rest are nonprofit organizations with varying degrees of affiliation with academic institutions. While this final selection was a result of availability and convenience, the results may have been quite different if the research focused only on national digital repositories, or on institutional repositories. Additionally, all but one of the repositories in this study are based in the United States. With a larger and more geographically diverse population, the study might have been able to examine regional or national trends in order to understand how nationality and/or location affect disaster planning activities.

6. Conclusion

This study found that while repositories are engaging in disaster planning activities, they are doing so largely as a means to obtain trusted digital repository status. Furthermore, repositories are reluctant or unwilling to share their disaster planning documentation. This

suggests that while one of the key elements of certification programs for digital repositories is the creation of formalized documentation of policies and procedures, these are not benefitting the community as much as they could. Since transparency is a core tenet of TRAC, auditors should insist that trusted digital repositories share disaster planning documentation and make non-sensitive policies and procedures available to the public in order to meet the criteria for trusted repository status, or to include the repository's documentation in the final audit report demonstrating that they have met the criteria for certification.

None of the repositories included in this study have had the opportunity to use their disaster planning documentation. While one hopes that these organizations will never have the need for their use, an opportunity for future research exists in the implementation and use of these documents. In her article on disaster preparedness, Schmidt observes that, "given enough time, the likelihood of a major disaster at an institution becomes a near certainty" (Schmidt, 2010).

7. Acknowledgements

I would like to thank Dr. Elizabeth Yakel, Dr. Paul Conway, and Shannon Zachary for their support and guidance throughout the course of this project.

References

Aikin, J. (2007). Preparing for a National Emergency: The Committee on Conservation of Cultural Resources, 1939-1944. The Library Quarterly, 77(3), 257.

Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C. (2009). Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences. The American Archivist, 72(1), 170-184.

Anderson, C. (2005). Digital Preservation: Will Your Files Stand the Test of Time? Library Hi Tech News, 22(6), 9-10.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired Magazine, 16(7). Retrieved from Wired Magazine website: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Andrew, G. (2008). Business Continuity: Best Practices. eWeek, 25(33), 32.

Babbie, E. R. (2010). The Practice of Social Research. Belmont, CA: Wadsworth Cengage.

Ball, A. (2010). Review of the State of the Art of the Digital Curation of Research Data: University of Bath.

Barateiro, J. A., Goncalo; Freitas, Filipe; Borbinha, Jose. (2010). Designing Digital Preservation Solutions: A Risk Management-Based Approach. The International Journal of Digital Curation, 5(1), 4-17.

Battersby, R. (2005). Recovering from disaster: the loss of Edinburgh's AI Library. Library + Information Update, 4(3), 36-36-38.

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). Keeping Research Data Safe: JISC.

Berman, F. (2008). Got data?: a guide to data preservation in the information age. Communications of the ACM, 51(12), 50-56.

Berman, F., Lavoie, B., Ayris, P., Choudhury, G. S., Cohen, E., Courant, P. N., . . . Van Camp, A. (2010). Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information; Blue Ribbon Task Force on Sustainable Digital Preservation and Access Final Report.

Best Practices in Disaster Recovery Business Continuity Planning. (2008). Baseline.

Brennan, C., & O'Hara, E. (2002). Murphy was a librarian: a case study in how not to handle a systems crash. Computers in Libraries, 22(3), 10-10-12.

Center for Research Libraries. (2007). Ten Principles. Retrieved April 2012, from http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re

Center for Research Libraries. (2010). CRL Certification Report on Portico Audit Findings.

Center for Research Libraries. (2011). CRL Certification Report on the HathiTrust Digital Repository.

Center for Research Libraries. (2012a). CRL Certification Report on Chronopolis.

Center for Research Libraries. (2012b). Reports on Digital Archives and Repositories. Retrieved on April 18, 2012 from http://www.crl.edu/archiving-preservation/digital-archives/digital-archive-reports

Cervone, H. F. (2006). Disaster recovery and continuity planning for digital library systems. OCLC Systems & Services: International digital library perspectives, 22(3), 173.

Constantinescu, C., Parulkar, I., Harper, R., & Michalak, S. (2008). Silent Data Corruption - Myth or reality? International Conference on Dependable Systems and Networks.

Cousins, T. J. (2007). Devising Post-Disaster Continuity Plans that Meet Actual Recovery Needs. IEEE Technology and Society Magazine, 26(3), 13.

Data Intensive Cyber Environments Group. (2008). iRODS: integrated Rule Oriented Data System White Paper: University of North Carolina at Chapel Hill, University of California at San Diego.

Data Seal of Approval. (2012). About Data Seal of Approval. Retrieved April 18, 2012, from http://datasealofapproval.org/

Educopia Institute. (2012). The MetaArchive Cooperative. Retrieved April 18, 2012, from http://www.metaarchive.org/

Fletcher, A. M. (2006). No Point of Reference: A Hurricane of Medical Information Needs. Journal of Hospital Librarianship, 6(2), 1-14.

Gantz, J. F., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W., & Toncheva, A. (2011). The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011. Framingham, MA.

Garrett, J., Waters, D. J. (1996). Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. [S.l.]: Distributed by ERIC Clearinghouse.

HathiTrust. (2012a). About HathiTrust. Retrieved April 18, 2012, from http://www.hathitrust.org/about

HathiTrust. (2012b). HathiTrust Exectuive Committee. Retrieved 2012, from http://www.hathitrust.org/xcom

Heiser, J. (2011). Best Practices for Recovering Critical Data From Damaged Hard Drives and Other Physical Media: Gartner Research.

Hey, T. T., Anne. (2003). The data deluge: an e-Science perspective. In F. H. Berman, A.; Fox, G. (Ed.), Grid Computing - Making the Global Infrastructure a Reality (pp. 809-824): John Wiley & Sons, Ltd.

Hitchcock, S., Brody, T., Hey, J. M. N., & Carr, L. (2007). Digital Preservation Service Provider Models for Institutional Repositories: Towards Distributed Services. D-Lib Magazine, 13(5/6).

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. Reading, Mass.: Addison-Wesley Pub. Co.

IInstitute of Risk Management, Association of Insurance and Risk Managers & Public Risk Management Association. (2002). A Risk Management Standard (p. 14): Institute of Risk Management, Association of Insurance and Risk Managers & Public Risk Management Association.

IInter-University Consortium for Political and Social Research. (2011, 2011). About ICPSR. Retrieved April 18, 2012, 2012, from http://www.icpsr.umich.edu/icpsrweb/ICPSR/org/index.jsp

Innocenti, P., & Vullo, G. (2009). Assessing the Preservation of Institutional Repositories with DRAMBORA: Case Studies from the University of Glasgow. Bollettino AIB, 49(2), 139-158.

Inter-University Consortium for Political and Social Research. (2011). ICPSR Timeline. Retrieved April 18, 2012, from http://www.icpsr.umich.edu/icpsrweb/ICPSR/org/timeline.jsp

International Organization for Standardization. (2012). Space data and information transfer systems — Audit and certification of trustworthy digital repositories (ISO 16363). Switzerland: International Organization for Standardization.

Internet Archive. (2012). About The Internet Archive. Retrieved April 18, 2012, from http://archive.org/about/about.php

Maniatis, P., Roussopoulos, M., Giuli, T. J., Rosenthal, D. S. H., & Baker, M. (2005). The LOCKSS peer-to-peer digital preservation system. ACM Transactions on Computing Systems, 23(1), 2-50.

MATRIX. (2012). About MATRIX. Retrieved April 18, 2012, from http://www2.matrix.msu.edu/about/

McDonald, R. H., & Walters, T. O. (2010). Restoring Trust Relationships within the Framework of Collaborative. Journal of Digital Information, 11(1).

McHugh, A. R., Seamus; Innocenti, Perla; Ruusalepp, Raivo; Hofman, Hans. (2008). Bringng Self-assessment Home: Repository Profiling and Key Lines of Enquiry within DRAMBORA. The International Journal of Digital Curation, 3(2), 130-142.

McKnight, M. (2006). Health Sciences Librarians' Reference Services During a Disaster. Medical Reference Services Quarterly, 25(3), 1.

Minor, D. Sutton, D.; Kozbial, A.; Westbrook, B.; Burek, M.; Smorul, M.. (2010). Chronopolis Digital Preservation Network. The International Journal of Digital Curation, 5(1), 119-133.

Moore, R. (2008). Towards a Theory of Digital Preservation. The International Journal of Digital Curation, 3(1), 63-75.

Muir, A., & Shenton, S. (2002). If the Worst Happens: The use and effectiveness of disaster plans in libraries and archives. Library Management, 23(3), 115-123.

Myles, B. (2000). The impact of a library flood on computer operations. Computers in Libraries, 20(1), 44-44-46.

National Library of Australia. (2012a). Digital Preservation. Retrieved April 18, 2012, from http://www.nla.gov.au/preserve/digipres/

National Library of Australia. (2012b). Digital Preservation Policy. Retrieved April 18, 2012, from http://www.nla.gov.au/policy-and-planning/digital-preservation-policy

National Library of Australia. (2012c). Digital Preservation Directions Statement 2008 to 2012. Retrieved from April 18, 2012, from http://www.nla.gov.au/digital-preservation-directions-statement-2008-to-2012

Nollau, B. (2009). Disaster Recovery and Business Continuity. Journal of GXP Compliance, 13(3), 51.

Patel, M. C., Simon. (2007). A Study of Curation and Preservation Issues in the eCrystals Data Reository and Proposed Federation. eBank-UK Phase 3: WP4, 1-34.

Patkus, B. L., & Motylewski, K. (1993). Disaster Planning. Preservation Leaflets. Retrieved from Northeast Document Conservation Center website: http://www.nedcc.org/resources/leaflets/3Emergency_Management/03DisasterPlanning.php

Pennock, M. (2007). Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information. Library & Archives(1), 3.

Portico. (2012). About Portico. Retrieved April 18, 2012, from http://www.portico.org/digital-preservation/about-us

Rajasekar, A. (2010). iRODS primer integrated rule-oriented data system. San Rafael, Calif. (1537 Fourth Street, San Rafael, CA 94901 USA): Morgan & Claypool Publishers.

Robson, C. (1993). Real world research: a resource for social scientists and practitioner-researchers. Oxford, UK ; Cambridge, Mass., USA: Blackwell.

Rombouts, J., & Princic, A. (2010). Building a 'data repository' for heterogenous technical research communities through collaborations. Paper presented at the International Association of Scientific and Technological University Libraries, 31st Annual Conference. http://docs.lib.purdue.edu/iatul2010/conf/day2/10

Ross, S. M., Andrew. (2006). Preservation Pressure Points: Evaluating Diverse Evidence for Risk Management. Paper presented at the iPRES 2006, New York, NY.

San Diego Supercomputer Center. (2011b). Infrastructure. Retrieved April 18, 2012, from http://chronopolis.sdsc.edu/infrastructure/index.html

San Diego Supercomputer Center. (2011a). About Chronopolis. Retrieved April 18, 2012, from http://chronopolis.sdsc.edu/about/index.html

Schmidt, G. (2010). Web 2.0 for Disaster Response and Recovery. Journal of Web Librarianship, 4(4), 413-426.

Schroeder, B. G., Garth A. (2007). DIsk failures in the real world: What does an MTTF of 1,000,000 hours mean to you? Paper presented at the FAST'07: 5th USENIX Conference on File and Storage Technologies, San Jose, CA.

Schultz, M. (2010). MetaArchive Cooperative TRAC Audit Checklist. Atlanta, GA: Educopia Institute.

Sesink, L.; van Horik, R.; Harmsen, H. (2010). Data Seal of Approval: Quality Guidelines for Digital Research Data (2nd Edition ed.). The Hague: Data Archiving and Networked Services (DANS).

Skinner, K. W., Tyler. (2011). New Roles for New Times: Digital Curation for Preservation, Published by ARL. Washington, D.C: Association of Research Libraries.

Tennant, R. (2001). Coping with disasters. Library Journal, 126(19), 26.

The Data Seal of Approval Board. (2011). Implementation of the Data Seal of Approval: Inter-University Consortium for Political and Social Research: Data Seal of Approval.

Wheatman, V. (2001). Aftermath: Disaster Recovery Aftermath. Stamford, CT: Gartner Research.

Wheatman, V. S., Donna; Witty, Roberta J. (2001). Aftermath: Business Continuity Planning Aftermath. Stamford, CT: Gartner Research.

Wildemuth, B. M. (2009). Applications of social research methods to questions in information and library science. Westport, Conn.: Libraries Unlimited.

Wong, Y. L., & Green, R. (2006). Disaster Planning in Libraries. Journal of Access Services, 4(3/4), 71 - 82.

Appendix A: Consent to Participate in a Research Study Interview

Invitation to Participate in a Research Study

You are invited to be a part of a study exploring how digital repositories are engaging in disaster planning activities.

Description of Subject Involvement.

Interview questions will concern the nature of your organization's attitude toward disaster planning activities, the process by which your organization's disaster plan was created, specific elements of that plan, and how your organization has implemented/used the plan. The interview will last approximately one hour.

Benefits.

Although you may not directly benefit from this study, others may benefit because of the comparison of digital disaster planning processes and documents.

Risks and Discomforts.

There are minimal risks associated with this study. Due to the fact that so few people are involved in digital disaster planning, the researchers are aware that you may not be able to remain completely anonymous.

Confidentiality.

We plan to publish the results of this study. If we quote from you, you will be given the opportunity to review and approve the use of any quotations that could be attributed to you. There are some reasons why people other than the researchers may need to see information you provided as part of the study. This includes organizations responsible for making sure the research is done safely and properly, including the University of Michigan and government offices.

All research records will be kept confidential to the extent provided by federal, state, and local laws. Data from the study will be kept in a secure location. These data may be used again in future research studies. Your real name will not appear in notes, transcripts or audio file names. Photographs, with permission, may be used in publications. You will be given a copy of this document for your records and one copy will be kept with the research records.

Consent.

Interviewees will get no direct benefits from this research. With your consent, this session will be audio taped.

Your participation in this study is voluntary. Subsequent to your consent, you may refuse to answer specific questions, withdraw from the study at any time, or ask that information be removed from our data set. You may also ask questions concerning the study before, during, or after the study.

Contact Information.

If you have questions about this research you may contact Rebecca D. Frank, University of Michigan, School of Information, (248) 854-0319, frankrd@umich.edu or Elizabeth Yakel, University of Michigan, School of Information, (734) 763-3569, yakel@umich.edu, fax (734) 615 - 3587.

This study has been reviewed by the Institutional Review Board at the University of Michigan and granted "Not Regulated" status. If you have any questions about your rights as a research participant, or wish to obtain information, ask questions or discuss any concerns about this study with someone other than the researcher(s), please contact the University of Michigan Institutional Review Board Health Sciences and Behavioral Sciences, (734) 936-0933, or toll free (866) 936-0933 540 E. Liberty St., Suite 202 Ann Arbor, MI 48104-2210, irbhsbs@umich.edu. (For international calls include the US Calling Code 1 and the exit number for the country of origin XXX+1+734-936-0933.)

Consent.

By signing this document, you are agreeing to be in the study. You will be given a copy of this document for your records and one copy will be kept with the study records. Be sure that questions you have about the study have been answered and that you understand what you are being asked to do. You may contact the researcher if you think of a question later.

I agree to participate in the study.

_____
Printed Name

_____ _____
Signature Date

Appendix B: Questions for Semi-Structured Interview

1. What is the nature of your organization's attitude toward disaster planning for digital collections? What types of activities has your organization undertaken regarding disaster preparedness/business continuity/continuity of service/etc.?

2. Does your organization have a formal disaster plan?
   a. For how long have you had a formal disaster plan?
   b. What was the impetus for the creation of this plan?
   c. Who is responsible for the maintenance of the plan?

3. If not, why? What documentation do you use in lieu of a formal disaster plan?
   a. Does your organization intend to create a disaster plan?
   b. What has prevented you from developing a plan thus far?

4. Please describe the process by which your disaster plan was created.
   a. Please discuss any of the following resources used in the creation of your organization's plan: standards, other organizations, other documentation and/or models, original research, professional organizations.
   b. What departments/individuals were involved in the creation of your disaster plan? Administration, digital preservation, systems, facilities?
   c. Did you consult with any external organizations? Backup sites, vendor agreements, power company, internet service providers, etc.?
   d. How long did it take? Were any parts of the process particularly time-consuming or challenging?
   e. Did the plan require any formal approval in your organization? Please discuss.

5. Discussion of specific elements of the plan. (Questions will vary depending on the organization's plan/documentation).

6. Please describe the process by which your disaster plan is updated. Do you have a formal review schedule to update the plan? Who is involved in this process?
   a. Who is responsible for maintenance of the plan? For updating the plan?

7. Is your organization TRAC Certified? Data Seal of Approval? Any other certifications?
   a. What impact has this certification had on your organization's disaster planning efforts?
   b. Did you have a disaster plan prior to certification?

8. What obstacles did you encounter during the development of your disaster plan? How did you overcome these obstacles/difficulties?

9. Who has access to the disaster plan?
   a. Are different versions available (i.e. a version for the general public vs. a restricted-access version for staff)?
   b. How is this plan made available (online, hard copy, etc.)?
   c. How are staff members (and possibly others) made aware of the plan?

d. Does your organization conduct training regarding the plan?

10. Have you had occasion to use the plan?
    a. Please describe the event(s).
    b. What elements of the plan were most useful? Least useful?
    c. How did your organization access the plan? Which method of access was most useful?
    d. How did this event affect your organization's view of the plan?
    e. Did this event prompt changes to the plan?

11. Please discuss how disaster planning fits into your organization's budget.