Disaster Planning for Digital Repositories

Rebecca D. Frank
University of Michigan School of Information
105 S. State St.
Ann Arbor, MI 48109-1285
frankrd@umich.edu

Elizabeth Yakel
University of Michigan School of Information
105 S. State St.
Ann Arbor, MI 48109-1285
yakel@umich.edu

ABSTRACT
This study examines how digital repositories with a preservation mandate are engaging in disaster planning, particularly in relation to their pursuit of trusted digital repository status. For those that are engaging in disaster planning, the study examines the creation of formal disaster response and recovery plans. Findings indicate that the process of going through an audit for certification as a trusted repository provides the incentive needed for the creation of formalized disaster planning documentation, and that repositories struggle with making their documentation available. This study also finds several significant obstacles with regard to the creation of formal disaster planning documentation, including the efforts required to get buy-in from different functional areas within the organization, difficulty collaborating with the IT department, and the amount of time required for completion of the documentation.

Keywords
Disaster Response and Recovery Planning, Digital Repositories, Trustworthy Digital Repositories.

INTRODUCTION
Disaster response and recovery planning remains one of the most important components of a preservation program in digital repositories. It is also one of the least understood and transparent. The adoption of standards and models for preservation, such as ISO 16363: 2012, the Space Data and Information Transfer Systems — Audit and Certification of Trustworthy Digital Repositories (2012) and ISO 14721:2012, the Open Archival Information System (OAIS) model have helped to clarify and illuminate best practices for disaster response and recovery in the digital preservation community. However, our understanding of disaster planning for digital repositories remains limited, both in terms of what constitutes disaster planning activities as well as whether any best practices have emerged in planning for different types of risk.

ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada.

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). This tipping point, the point at which data created outpaced our capacity to store it, was significant for the digital preservation community. It was at this point when decision making for digital preservation needed to 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, "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, p. 16). The importance and uniqueness of research data, 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 (Gracy & Kahn, 2012), but a thorough understanding of disaster planning in practice has not yet been achieved.

This study is driven by the following research questions:

1. What motivates the disaster planning activities in digital repositories?
2. What is the scope of disaster plans for digital repositories?
3. Does the pursuit of trusted digital repository status affect disaster planning activities?

In this pilot study, we concentrate on the disaster planning 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. The study consists of 10 interviews with individuals from 8 different organizations. 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 (Consultative Committee for Space Data Systems, 2012; Innocenti & Vullo, 2009; McHugh, Ross, Innocenti, Ruusalepp, & Hoffman, 2008).

LITERATURE REVIEW

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 emerged in the analog era, but has become vital in the digital era (Aikin, 2007; Cervone, 2006; Muir & Shenton, 2002; Myles, 2000). Digital preservation consists of actions that ensure the viability and authenticity of digital objects over time and disaster planning is one of those actions (Berman, 2008). 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). More specifically, a disaster plan is a document that describes policies and procedures that have been created to prevent, prepare for, respond to, and recover from a disaster (Muir & Shenton, 2002). Beyond this, nothing is written on disaster planning for digital repositories. A majority of the evidence we have for this activity are copies of digital disaster plans on the web (e.g., Inter-University Consortium for Political and Social Research (ICPSR), 2013).

Some approaches to digital preservation, such as the LOCKSS (Lots of Copies Keeps Stuff Safe) system, have implicit disaster planning strategies built in. 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 duplication across many different sites. 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 (Maniatis, Roussopoulos, Giuli, Rosenthal, & Baker, 2005). In a study published in 2007, Schroeder and Gibson (2007) conducted a survey that suggested that the random storage faults discussed by Maniatis et al. are indeed likely to occur. However, these articles do not focus specifically on what types of disaster responses are needed to guard against these faults or how recovery planning should occur when they are detected. Instead, these two articles imply that disaster response and recovery planning are not needed – if implemented properly the duplication for long-term preservation will allow the system to overcome any type of disruption or loss in service.

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, 2008). The iRODS system is not specifically a preservation system, but it can be used to facilitate and support preservation by mitigating against the risk of disaster because it allows repositories to create and enforce rules and policies, therefore ensuring consistency within the repository (Rajasekar et al., 2010).

Digital Curation

Digital curation involves assessing value for collections and implies some type of value proposition. According to Walters and Skinner (2007), "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" (Skinner, 2006). Susceptibility to disasters is a problem not only if it interrupts access to collections, but also if it threatens the integrity of those valued collections, whether they are needed for one year or twenty.

Trust

Another important element of preservation and disaster preparedness for digital repositories is the concept of trust. Garrett and Waters (1996) 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" (Garrett & Waters, 1996).

The concept of trust has emerged in community standards for digital repositories. Through the granting of trusted digital repository status, repositories are deemed trustworthy, and part of that designation involves evidence of disaster planning and response activities and documentation. Several mechanisms through which repositories can gain trusted status are: Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist (ISO 16363); Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) which was developed jointly by the Digital Curation Centre and DigitalPreservation Europe; Data Seal of Approval (DSA), which originated in the Netherlands, and Network of Expertise in Long-Term Storage and Long-Term availability of Digital Resources (nestor), the German competence network for digital preservation.

Disaster planning is a core construct of the TRAC requirements. Sections 5.1 and 5.2 are most explicit, "the repository shall have suitable written disaster preparedness and recovery plan(s)" (Consultative Committee for Space Data Systems, 2012) TRAC certification requires that the repository be able to demonstrate disaster preparedness. This preparedness is generally demonstrated through the creation of a disaster plan or, more accurately, a suite of disaster planning documents. In fact, two of the four certified repositories, Chronopolis and HathiTrust, had to specifically create disaster plans to gain certification.

The focus on risk in the DRAMBORA assessment can arguably be seen as analogous to the TRAC requirement for disaster preparedness as both seek to mitigate the risk that disasters pose to digital repositories. The DRAMBORA assessment, in fact, has a stronger focus on risk management and mitigation as the entire assessment is based on a repository's ability to assess, manage, and respond to risks. In both cases, there is a strong emphasis on preparation for and the ability to respond to disasters.

The 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). While disaster planning and risk management are not explicitly discussed, the focus on digital archiving, long-term preservation, and lifecycle management are relevant to disaster planning.

The German nestor project is a competence network that has produced a catalog of criteria for trusted digital repositories (nestor Working Group Trusted Repositories - Certification, 2009). This listing of criteria specifies that repositories should be prepared to address and respond to risks, but does not specifically state that the repository should have a disaster plan.

Key differences exist between TRAC, DSA, nestor, and DRAMBORA: TRAC and DSA provide strict guidelines for performing an audit; nestor provides a listing of criteria for trusted repositories; and DRAMBORA provides a framework that can be adapted to fit the needs of any repository (Ball, 2010; Center for Research Libraries, 2007; Data Archiving and Networked Services (DANS), 2010; Dobratz, Schoger, & Strathmann, 2007; nestor Working Group Trusted Repositories - Certification, 2009; Patel, 2007). Despite differences in philosophy and degree of formality, TRAC, DSA, nestor, and DRAMBORA all specifically include requirements for repositories to have disaster planning and risk management documentation (Data Archiving and Networked Services (DANS), 2010; McHugh et al., 2008; Ross & McHugh, 2006)

Planning for Disasters

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

> *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 (Altman et al., 2009).

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, there are some types of disasters that are outside of the control or influence of the repository and for these types of disasters repositories must do what they can to mitigate data loss (Anderson, 2005; McDonald & Walters, 2010). In anticipation of the need to recover from data loss, repositories are moving toward the widespread adoption of best practices for preservation. These best practices also contribute to the granting of trusted repository status as described above with TRAC, DSA, nestor, and DRAMBORA certifications.

Literature discussing disaster planning for digital repositories is sparse, and as such discussion is necessarily limited (Anderson, 2005). However, the general trends discussed above, and the recognition by the community that disaster planning is a beneficial and recommended action for digital repositories, are promising and suggest that this is an area that will continue to expand.

METHODOLOGY

The methodology of this study involves 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.

Selection of Sites

The sample population for this study consisted of digital repositories that have taken some steps toward 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. The initial list of 19 repositories was created in May of 2011, based on information available at that time. Eight were selected for inclusion in the final study based on their availability and the willingness of individuals at those organizations to be interviewed. All who were able to complete an interview by the end of January 2012 were included in the study.

Interviews

Interview subjects were identified at each of the initial 19 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. Individuals at these organizations 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, and ten individuals were ultimately selected for participation.

Of the ten subjects interviewed for the study, three hold administrative roles, five digital preservation roles, and two 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 their respective organization.

Participants were sent consent forms to review and sign prior to the interview. Interviews lasted approximately one hour each and were conducted via telephone or in person. All interviews were recorded and transcribed.

The interviews followed a semi-structured list of questions, which allowed for modification to the wording and/or order of the questions. This also allowed for further probing and requests for elaboration to clarify the subjects' responses where necessary (Babbie, 2010; Robson, 1993; Wildemuth, 2009). Questions covered the areas of: organizational attitudes toward disaster response and recovery planning, development of, access to, use, and maintenance of disaster planning documentation, and budgetary considerations.

Code	Repository	Role
Subject A	Chronopolis	Digital Preservation
Subject B	Chronopolis	Digital Preservation
Subject C	HathiTrust	IT
Subject D	HathiTrust	Digital Preservation
Subject E	ICPSR	Digital Preservation
Subject F	MATRIX	IT
Subject G	National Library of Australia	Digital Preservation
Subject H	Portico	Administration
Subject I	The Internet Archive	Administration
Subject J	The MetaArchive Cooperative	Administration

Table 1: Repositories and Interview Subjects.

Interview Analysis

Once completed, the interviews were transcribed and coded. The system for coding 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 (Holsti, 1969; Wildemuth, 2009). Thus, the coding scheme was both grounded or bottom up, arising from the interviews themselves and top down based on themes in the literature. Codes fell into general categories of communication, documentation, administration, and preservation. In all instances, the first round of coding was descriptive. More interpretative and pattern analysis was done in a second round of examination once the initial coding has been completed.

Limitations

There are several 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 have engaged in trusted repository audits. 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. Third, the sample did not represent diverse repositories. 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.

FINDINGS

The coded interview data yields findings in the following three areas:

1. Incentives for Creation of a Disaster Response and Recovery Plan
2. Disaster Response and Recovery Plan Documentation
3. Process of Creation of a Disaster Response and Recovery Plan, Including Obstacles

Incentive for Creation of a Disaster Response and Recovery Plan

Many of the subjects claimed that disaster planning activities occurred as a result of the organization's growth and development. However, even those who insisted that their repository had disaster planning policies and procedures in place indicated that it was in response to the requirements of the certification process that they created formal disaster response and recovery planning documents.

Most of the interviewees from organizations that have received certification as a trustworthy digital repository (i.e. Chronopolis, HathiTrust, and Portico) specifically stated that formal disaster planning documentation had been created for the audit. As Subject A from Chronopolis stated, "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." Interviewees from HathiTrust indicated that they were in the process of creating formal disaster planning documentation as required by their previous TRAC audit, and expected to have it completed for their next certification review.

Similarly, Subject J from The MetaArchive Cooperative 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."

Staff from the repositories without official certification claimed to have disaster planning policies and procedures in place. Yet, they did not have formal disaster planning documentation. The Internet Archive, MATRIX, and the National Library of Australia fell into this category. Interviewees from each of these organizations explained that while policies and procedures existed, they had not found the need to create formal disaster planning documentation. 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."

Among the repositories included in this study, ICPSR was the only organization to create a full suite of formal disaster planning documentation independent of any audit or certification process, although they later applied for and received a Data Seal of Approval. Subject E explained that the development of disaster response and recovery policies and procedures was a result of organizational growth and development, "I think it was just a general sense of alignment with good practice . . . there was a sense that we need some kind of disaster planning in place."

Disaster Response and Recovery Plan Documentation

All of the interviewees stated that their organization had some form of disaster planning documentation. Many provided evidence of that documentation in the form of an audit report, but only ICPSR and HathiTrust were able or willing to discuss these plans in detail or to provide copies of their complete documentation. This is significant in part because one of the key tenets of trustworthy digital repositories is transparency (Consultative Committee for Space Data Systems, 2012). A repository that does not make key documentation available for review is arguably not transparent. That said, we acknowledge that some disaster planning documentation needs to be confidential for security reasons.

Both interviewees from Chronopolis explained that the organization's disaster planning documentation was created in response to the recent TRAC audit. "In general terms we created a TRAC report which basically follows the question and answer schema of the TRAC audit itself" (Subject A). 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, "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 our 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 partner institution.

In an interesting exchange, Subject F (a technologist from MATRIX) 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 has created several different documents that comprise their disaster planning documentation, "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" (Subject H). 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 seemed to be describing a disconnect between the IT and Preservation functions within the organization. Rather than having one set of combined disaster planning documents, Portico had two separate sets of documents that were not combined in any formal or significant way. In fact, the interviewee was able to discuss the IT documents in only broad strokes.

The interviewees from 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 were in place, and that the creation of formal disaster planning documentation was 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 interviewee from ICPSR was also completely open about sharing her disaster documentation. "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" (Subject E). Nearly all ICPSR 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 (IA) have, "an internal checklist absolutely which we review" that is maintained by an IT department. However, IA staff do not consider the checklist to be a complete disaster response and recovery plan. It is also not available to the public. 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 was so great that the digital preservation section staff were not familiar with the disaster planning documentation at all.

Process of Creation of a Disaster Response and Recovery Plan, Including 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. Common themes included organizational obstacles, such as the difficulty of getting buy-in from other internal members of the organization, in particular difficulty collaborating and communicating with the IT department, and the amount of time required for completion of the documentation. These obstacles align with the earlier finding that most repositories with formalized disaster planning documentation created that documentation as the result of an audit.

Several of the interviewees discussed the difficulty of creating disaster planning documentation while experiencing resistance from other members of their organizations. For example, Subject A 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 . . . in order to get detailed documents it really did take the audit to pull those things out."

Subject E 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 or that of their subordinates 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."

Collaboration with the IT department was an obstacle identified by several interviewees. Subject G described a great deal of difficulty in working with the IT section of the organization, including difficulty convincing them that digital preservation and disaster planning are valid or necessary. Conversely, Subject F explained the decision not to create formal disaster planning documentation in the following way, "most System Administrators 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." These two sides of the same issue illustrate the difference in perspective between the preservation and IT functions within these repositories.

Subject E from ICPSR described a series of drills that the repository staff ran in order to test their disaster recovery capabilities. They found that they were able to recover a full copy of the repository's data from backup, but that it was more difficult and took significantly more time than they had expected. "We have done a complete – we have six copies online and then one copy on a tape – and we have actually, as part of … the fire drill we did and it actually took longer than they thought which was exactly why I wanted to do it but they were actually able to reconstitute all of the stored files from the tape backup to demonstrate that they could."

The amount of time needed to produce formal disaster planning documentation was identified as another significant obstacle. Interviewees from organizations both with and without formal documentation in place discussed this challenge extensively. "I'm interested in seeing the cost/benefit analysis of how much time does it take, how much effort went into the creation of 'x' part of the document vs. how useful that part of the document would even be in terms of disaster recovery process" (Subject D).

All of the repositories that have been through some type of audit described the process as being time consuming. According to Subject H, "I would say that it was probably a six month process for us . . . 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." For Chronopolis, Subject A explained that the creation of formal disaster planning 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 process of creating formal disaster planning documentation was described by most subjects as time consuming and difficult, despite the claim by most interviewees in this study that creating formal disaster planning documentation was merely a process of documenting current practices. For The MetaArchive Cooperative, "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."

DISCUSSION

In this section, we discuss the three main findings of this study in greater detail. The three major findings of this study address:

1. The incentive that certification provides for disaster planning.
2. The lack of transparency in digital disaster planning.
3. The need for Coordination between the IT and preservation functions.

For most organizations, the process of going through an audit for certification as a trusted repository provided diverse staff with an incentive to allocate time to create formalized disaster planning documentation. Organizations that have been through an audit process were more likely to have formal disaster planning documentation in place. Interviewees from these organizations discussed the role that the audit played in providing widespread organizational 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, such as the need for certification, it was difficult or impossible to convince staff in other functional

areas, such as IT and Administration, to spend time documenting policies and procedures that were not formally documented or perhaps even articulated elsewhere or were just tacitly understood.

These findings suggest that one of the primary benefits of achieving trusted digital repository status, in addition to certification itself, is the fact that it provides an opportunity for organizational members to enter into a dialog about disaster prevention and recovery and an incentive for the entire organization to create accurate, up-to-date, thorough documentation of policies and procedures. For organizations that already have some form of documentation in place, the audit provides the organization with an opportunity to improve and update their documentation.

Repositories struggle with the decision to make disaster planning documentation available to the general public. We expected that repositories that had been through an audit for certification would be willing to make at least parts of their disaster planning documentation publicly available, perhaps with some restrictions due to sensitive security-related information. Given that transparency is one of the value principles of TRAC, it seemed natural that TRAC certified repositories would then make at least parts of their documentation available to the public. As we discovered, this was generally not the case. The availability of documentation regarding disaster planning activities varied widely among repositories and ran the full spectrum from fully available to completely restricted.

Finally, we found that the single greatest obstacle to disaster planning activities at all stages of the process was 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. IT staff viewed their tacit understanding and procedures as sufficient. Other staff wanted greater articulation of those procedures and transparency in policies.

We also saw 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. Those in the field of digital preservation need to find ways of communicating with those in IT in order to improve collaboration and coordination throughout the organization to address the management of digital assets.

The initial research question for this study focused on investigating how repositories engage 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 have formal documentation regarding their disaster planning efforts. 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 full disaster planning documentation publicly available was a major limitation for this study. Additionally, this lack of transparency may be hampering disaster planning efforts in the community as few models of the documentation or the disaster planning process are available for review.

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 some 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. In an 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). While one hopes that these organizations will never have the need for their use, this suggests that they will, and an opportunity for future research exists in the implementation and use of these documents.

ACKNOWLEDGMENTS

We would like to thank Dr. Paul Conway and Shannon Zachary for their comments on earlier drafts of this paper.

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.

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

Ball, A. (2010). *Review of the State of the Art of the Digital Curation of Research Data* (ERIM Project Document erim1rep091103ab11 No. version 1.1) (p. 56). Bath, UK: University of Bath. Retrieved from http://opus.bath.ac.uk/18774/2/erim1rep091103ab11.pdf

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe*. JISC. Retrieved from http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx

Berman, F. (2008). Got Data?: A Guide to Data Preservation in the Information Age. *Communications of the ACM*, *51*(12), 50–56. doi:10.1145/1409360.1409376

Center for Research Libraries. (2007). Ten Principles. *Center for Research Libraries Global Resources Network*. Retrieved October 9, 2012, from http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re

Cervone, H. F. (2006). Disaster Recovery and Continuity Planning for Digital Library Systems. *OCLC Systems & Services: International Digital Library Perspectives*, *22*(3), 173.

Consultative Committee for Space Data Systems. (2012). *Space Data and Information Transfer Systems — Audit and Certification of Trustworthy Digital Repositories* (Standard No. ISO 16363:2012 (CCSDS 652-R-1)). Washington, D.C.: Consultative Committee for Space Data Systems.

Consultative Committee for Space Data Systems. (2012). *Reference Model for and Open Archival Information System (OAIS)* (Magenta Book No. CCSDS 650.0-M-2). Washington, D.C.: Consultative Committee for Space Data Systems.

Data Archiving and Networked Services (DANS). (2010). *Data Seal of Approval: Quality Guidelines for Digital Research Data*. The Hague. Retrieved from http://www.datasealofapproval.org/?q=about

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

Dobratz, S., Schoger, A., & Strathmann, S. (2007). The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification. *The Journal of Digital Information*, *8*(2). Retrieved from

Garrett, J., & Waters, D. J. (1996). *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (No. 9781887334501 1887334505) (p. 68). Washington, D.C.: The Commission on Preservation and Access & Research Libraries Group.

Gracy, K. F., & Kahn, M. B. (2012). Preservation in the Digital Age. *Library Resources & Technical Services*, *56*(1), 25–43.

Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley Pub. Co.

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 (ICPSR). (2013). Disaster Planning. *ICPSR Data Management: Disaster Planning*. Retrieved April 1, 2013, from http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/disaster/

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. doi:10.1145/1047915.1047917

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

McHugh, A., Ross, S., Innocenti, P., Ruusalepp, R., & Hoffman, H. (2008). Bringng Self-Assessment Home: Repository Profiling and Key Lines of Enquiry within DRAMBORA. *The International Journal of Digital Curation*, *3*(2), 130–142. doi:doi:10.2218/ijdc.v3i2.64

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.

Nestor Working Group Trusted Repositories - Certification. (2009). *nestor Criteria: Catalogue of Criteria for Trusted Digital Repositories, Version 2*. Frankfurt am Main: Deutsche Nationalbibliothek.

Patel, M. C. (2007). *A Study of Curation and Preservation Issues in the eCrystals Data Reository and Proposed Federation* (Final Version (Revised) No. eBank-UK Phase 3: WP4) (pp. 1–34). Bath: UKOLN, DCC; National Crystallography Centre.

Rajasekar, A., Moore, R., Hou, C.-Y., Lee, C. A., Marciano, R., de Torcy, A., … Zhu, B. (2010). iRODS Primer: Integrated Rule-Oriented Data System. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *2*(1), 1–143. doi:10.2200/S00233ED1V01Y200912ICR012

Robson, C. (1993). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Oxford, UK ; Cambridge, Mass., USA: Blackwell.

Ross, S., & McHugh, A. (2006). Preservation Pressure Points: Evaluating Diverse Evidence for Risk Management. Presented at the iPRES 2006, New York, NY: Digital Curation Centre.

Schmidt, G. (2010). Web 2.0 for Disaster Response and Recovery. *Journal of Web Librarianship*, *4*(4), 413–426. doi:10.1080/19322909.2010.511038

Skinner, R. E. (2006). "Nor Any Drop to Drink": New Orleans Libraries in the Aftermath of Hurricane Katrina. *Public Library Quarterly*, *25*(3-4), 179–187. doi:10.1300/J118v25n03_15

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.