

Margaret L. Tuomala. "That's a Good Question...": Asking Information Professionals About Digital Preservation at a University's Digital Institutional Repository. A Master's Paper for the M.S. in L.S. degree. July, 2010. 38 pages. Advisor: Cal Lee

This study addresses some of the issues surrounding the preservation of digital objects within a digital repository. The issues of how to determine the significant properties of digital objects and how digital repository infrastructures are developed to foster preservation are specifically addressed. Semi-structured interviews with information professionals and researchers working with a university's digital institutional repository were conducted to gather professional opinions on these issues. Results show the need for a clearly defined goal, concrete policies, and a sustainable repository model as some of the most important factors when creating a trustworthy digital repository.

Headings:

Institutional repositories.

Digital preservation.

Electronic archives.

College and university libraries.

Interprofessional cooperation.

“THAT’S GOOD QUESTION...”: ASKING INFORMATION PROFESSIONALS  
ABOUT DIGITAL PRESERVATION AT A UNIVERSITY’S DIGITAL  
INSTITUTIONAL REPOSITORY

by  
Margaret L. Tuomala

A Master’s paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Library Science.

Chapel Hill, North Carolina

July 2010

Approved by

---

Cal Lee

As defined by the Joint Information Systems Committee (JISC),<sup>1</sup> digital preservation is “the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value... [and] encompasses not just technical activities, but also all of the strategic and organizational considerations that relate to the survival and management of digital material” (2006). Presently, digital collections and digital objects within collections are becoming an increasingly common, even expected, feature of libraries and archives. In the archival and information and library science communities digital preservation is a very current topic of research, most of which focuses on establishing best-practices, recommendations, and guidelines for implementing and sustaining digital preservation projects. When we consider that the aim of digital preservation, according to the “Significant Properties Report” by Andrew Wilson, is to ensure “that authentic digital objects remain accessible and usable over time,” and acknowledge that in order to responsibly preserve digital objects it is necessary to understand what their significant properties are, we come to the conclusion that this specific focus of research under the larger scope of digital preservation is essential to the understanding of digital preservation (2007, p. 3). However, this assertion raises some problems. First, it is often difficult to discern exactly what the significant properties of a digital object are; second, they frequently change depending on who is accessing and using the digital object; and

third, the myriad significant properties of digital objects are not always explicitly presented to the information professionals, archivists, and librarians involved in digital preservation.

### **Problems in Digital Preservation and Research Questions**

As stated above, it is often difficult to discern exactly what the significant properties of a digital object are, and they can change depending on who is using the digital object. Additional problems arise when the myriad significant properties of digital objects are not explicitly presented to the information professionals involved in digital preservation.

Currently, a research-oriented public university is developing its own institutional repository for digital objects. This presents a unique opportunity to research digital preservation on an institutional level at a specific digital institutional repository. This study asks the following research questions: what specific steps can the information professionals involved in digital preservation at the university's digital institutional repository take to make sure that these properties are preserved for the long-term, and how is the university's digital institutional repository infrastructure fostering digital preservation?

### **Literature Review**

This literature review will explore digital preservation research and delve into the literature that describes digital objects and begins to identify, or establish criteria for identifying, their most significant properties. First, some of the terms that are most prevalent in and specific to the literature will be defined. Then the review will move on to an overview and discussion of the projects and the studies that currently inform

research on digital preservation. Although most of the research presented in this literature review directly addresses the topic of significant properties of digital objects, some speaks of institutional readiness in regards to digital preservation, and even the broader, somewhat philosophical, implications that digital preservation has on the field of archival science.

### **Definitions Specific to Research in Digital Preservation**

It is useful to determine a solid and established definition of what the literature refers to as a digital object. As defined by Stephen L. Abrams in 2005

in general, a digital object can be considered as the encapsulation in digital form of some piece of abstract intellectual content. More specifically, a digital object is the aggregation of one or more formatted content streams representing the primary content of the object as well as associated descriptive, administrative, technical, and structural metadata. (p. 125)

Using this broad definition the literature refers to and considers a digital object to be both born-digital content as well as previously analog content that has been reformatted into digital form.

It is also helpful to define what the literature speaks of as the significant properties of digital objects. As set forth by Margaret Hedstrom and Christopher A. Lee in a 2002 paper, “significant properties are those properties of digital objects that affect their quality, usability, rendering, and behavior” (Berninger, Brady, Hofmann, & Schram, eds., p. 218). Another workable and expressive definition for what Hedstrom and Lee term significant properties is presented by Helen Heslop, Simon Davis, and Andrew Wilson who refer to the “essence of a record” and define this as “those characteristics of the context, rendition, and structure of a digital record that must be preserved together with the content to give the record meaning” (2002, p. 7). Perhaps the most widely

recognized definition of significant properties is presented by Andrew Wilson who states that significant properties are “the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects” (2007, p. 8). It may also be important to note that some of the literature refers to these as “significant characteristics” (Agosti, Borbinha, Kapidakis, Papatheodorou, & Tsakonas, 2009, p. 297). However, in keeping with the popular terminology of digital preservation this literature review will use the term significant properties.

### **Digital Preservation Projects**

There have been several projects exploring digital objects and their significant properties. Findings from these projects have allowed the archival, library and information science, and digital preservation communities to commence the exploration of significant properties, to start to make decisions about just what they are, and to investigate, in turn, how they may begin to be preserved in digital repositories.

The Cedars<sup>2</sup> Project, spanning from 1998 to 2001, sought to “develop a demonstrator system to recommend techniques for long term storage of digital data primarily within the research library context,” while conforming to the Open Archival Information System (OAIS) model<sup>3</sup> (Holdsworth & Sergeant, 2000, p. 1). Established by the Consortium of University Research Libraries (CURL) and funded by the JISC Electronic Libraries (eLib) program, the Cedars Project raised awareness of and addressed the practical issues surrounding digital preservation and provided guidance in best practice for digital preservation, as was the source of the concept of significant properties (The Cedars Project Team, 2001). The Cedars Project introduced the idea that an interaction with a digital object should include a recreation of the experience of

viewing and even interacting with the original, an idea that is elaborated by David Holdsworth and Derek M. Sergeant. (2000, p. 3). Holdsworth and Sergeant highlight that such an experience relies on a “recreation of the significant properties of the original digital object” and assert that ingest standards should require that these significant properties be identified (2000, p. 3).

Another project relevant to the study of the significant properties of digital objects is the Creative Archiving at Michigan and Leeds: Emulating the Old on the New (CAMiLEON) project, which overlapped in both time and personnel with the Cedars Project. The CAMiLEON project spanned from 1999 to 2003 and sought to

test the feasibility of emulation as a digital preservation strategy, to evaluate its effectiveness in preserving the original ‘look and feel’ and behavior of various types of complex digital objects, and to define the attributes of different types of digital objects that must be preserved to satisfy user needs and requirements. (Hedstrom & Lampe, 2001, Introduction, para. 2)

The CAMiLEON project is especially important because it produced a paper authored by Margaret L. Hedstrom, Christopher A. Lee, Judith S. Olson, and Clifford A. Lampe that presented the findings of a study that investigated digital preservation from the user’s perspective. This study and resulting paper will be discussed in further detail in the next section of this literature review.

Following the Cedars and CAMiLEON projects, the most recent major initiative related to the “development of a formal, or canonical method to define significant properties” was InSPECT,<sup>4</sup> a project performed by the Arts and Humanities Data Service and the National Archives in the UK, and funded by JISC (Knight, 2008, p. 3). The InSPECT project sought to

examine the whole concept of significant properties, determine which properties are significant for a range of object types and assess the importance of each of

these for future representation of the object, and finally propose a generalized methodology that will enable resource curators to determine the significant properties of classes of digital objects that must be preserved over time. (2007, Aims and Objectives, para. 1)

Two outputs of the InSPECT project, papers by Andrew Wilson and Gareth Knight, reflect many of the recent movements in digital preservation. These papers build upon previous projects and studies and provide an extremely clear and lucid articulation of the concept of the significant properties of digital objects, and “outline a methodology for the identification and description of significant properties contained by a digital resource” (Wilson, 2007, p. 2; Knight, 2008, p. 3).

### **Digital Preservation Studies**

Several studies, especially those coming out of the projects presented above, have begun to address the importance of the preservation of significant properties of digital objects; however, it is clear that there are some gaps in the current body of research on digital preservation. This literature review will present and discuss four studies that address some of the current issues within the field of digital preservation. Two of these studies deal directly with the questions surrounding significant properties from the user’s perspective, while the remaining two address the organizational and curatorial questions raised by studies in digital preservation.

In their 2006 article, produced as a result of the CAMiLEON project, Margaret L. Hedstrom, Christopher A. Lee, Judith S. Olson, and Clifford A. Lampe discuss digital preservation from the user’s point of view, aiming to “understand which features users consider worth preserving, rather than what archivists believe is important or what theoretical models would predict” (p. 160). Hedstrom et al. conducted “two laboratory experiments with human subjects,” observing as these subjects “interacted with digital



materials,” and gathering “comments from them about the appearance and behavior of digital materials” (2006, p. 166). The digital materials that the subjects interacted with in these experiments were original, emulated, and migrated “versions of an obsolete computer game called ‘Chuckie Egg’” and different formats or “versions of speech files and office documents from a former president of the University of Michigan” (Hedstrom et al., 2006, p. 166). The experiments using Chuckie Egg found that “the subjects noticed very minor differences” between the versions of the game, and that “subjects preferred playing the migrated and emulated versions rather than the original game” (Hedstrom et al., 2006, p. 171). The experiments using the different formats of speech files and office documents found that while “subjects used a complex reasoning process... to judge the authenticity of digital documents” they often made incorrect assumptions that caused them to make errors in their judgments of these digital documents, and “only one subject used any of the metadata associated with [the] documents to help determine which was the original” (Hedstrom et al., 2006, p. 183-184). They also found that “usability is the primary factor that will influence future users’ preferences for the document format they would choose in a research project and that authenticity is at best a secondary consideration” (Hedstrom et al., 2006, p. 184). The combined findings from the Chuckie Egg and University of Michigan president’s documents experiments address the “significance of ‘look and feel’ for preserved digital objects and the importance of contextual information in helping users understand and interpret digital information” (Hedstrom et al., 2006, p. 185).

In a 2008 article<sup>5</sup> Kellie Snow et al. state that the purpose of their research is to “deliver a deep understanding and appreciation of digital object creation and use as well

as how these processes may impact on digital preservation” (Research Focus, para. 1). Specifically, they are “concentrating on the usage of digital objects in libraries, archives, and data centers, investigating the issues that are important to users when working with this information” (Snow et. al 2008, Research Focus, para. 2). Using the Contextual Design process, a user-centered process developed of understanding how the user works, augmented with other social science data collection techniques as the framework for gathering information from their subjects, Snow et al. conducted a usage study on users working in libraries, data centers, and archives over a period of two to four weeks (Beyer & Holtzblatt, 1998, p.21). They then analyzed the information gathered, discovering “a number of themes that are significant to users when working with digital information...: authenticity, provenance, access and usability, search and selection, and granularity of data” (Snow et al. 2008, Findings, para. 1). Like the Hedstrom et al. article, this article addresses the issue of significant properties of digital objects as seen through the eyes of the user.

Another article that informs the research surrounding digital preservation is Daniel G. Dorner’s “Public Sector Readiness for Digital Preservation in New Zealand: The Rate of Adoption of an Innovation in Records Management Practices” (2009). The research presented in this article “identifies the level of organizational awareness” of some of the issues and risks surrounding digital preservation, such as “technological obsolescence, lack of organizational policies, insufficient resources, fragile storage media, and other threats” as well as of “the current activities taking place in this area in order to determine the present state of readiness for digital preservation in New Zealand’s public sector” (Dorner, 2009, p. 341). In order to conduct this study, Dorner first

identified his research question as “what is the current state of readiness for digital preservation within the New Zealand public sector?” (2009, p. 342). He then developed a web-based survey which was “conducted of the nation’s government organizations in March 2006” and “achieved a response rate of 42.4 percent” (Dorner, 2009, p. 341). Dorner’s survey “found that most organizations were knowledgeable about basic aspects of their digital resources but their awareness of digital preservation was generally low” and they “are adopting digital preservation at a relatively slow rate” (2009, p. 341, 346). The survey also found that “digital preservation activity” at these government organizations “could best be described as modest overall” (Dorner, 2009, p. 346). With these findings in mind, Dorner asserts that “the organizations are not sufficiently aware of the positive advantages that will accrue to them in terms of their ongoing access to digital records” and that “in order for the rate of adoption of digital preservation to increase, it will need strong champions” who can communicate and promote the innovations of digital preservation clearly to the custodians and creators of digital records (2009, p. 347).

Although Daniel G. Dorner’s 2009 article does not explicitly address the idea of significant properties, it is relevant to the topic of digital objects and significant properties because of its take on the organizational side of digital preservation, and the picture it paints of institutional readiness. While the Hedstrom et al. and Snow et al. articles previously discussed speak of users and their needs, it is also important to understand the needs of an institution and its professional staff. In this article, Dorner builds a frame that can be used for further studies in how institutions and staff working

with the preservation of digital objects can prepare to best preserve the significant properties of digital objects.

Susan Thomas and Janette Martin's 2006 article is a product of the Personal Archives Accessible in Digital Media (Paradigm) project. Paradigm is "an exemplar project... exploring the cultural, legal and technical issues involved in the long-term preservation of digital private papers by engaging with record creators and employing sample collections to practice archiving digital private papers" (Thomas & Martin, 2006, p. 36). Specifically, the Paradigm project is "using the papers of contemporary British politicians as a testbed" for looking at "how archivists might select, acquire, process, store, preserve, and provide access to the digital archives of individuals for the use of future researchers" with the "goal of striking a balance between theoretical principles and practical procedures" in archival science, digital curation, and digital preservation (Thomas & Martin, 2006, p. 29). Thomas and Martin explain that while there has been previous research on the management and preservation of digital objects, most of this has "been geared towards developing workflows, standards, and systems capable of creating, acquiring, enhancing, storing, and retrieving digital [objects]" and that less attention has been paid to "developing preservation functions for such systems" (2006, p. 31). They also cite growing interest in the "preservation aspect of digital curation," bringing another aspect to digital preservation to the table, and they address provenance and context as significant properties especially important to their project (Thomas & Martin, 2006, p. 31, 41).

Like the Dorner article discussed above, Thomas and Martin do not explicitly address the preservation of significant properties. However, they do speak of the need to

involve the creators of digital objects in the digital preservation process. This idea is both novel and useful in that it suggests that the creators of digital content will be able to most clearly define the significant properties of their personal digital objects.

## **Conclusion**

The literature presented above provides a solid foundation for further research into the significant properties of digital objects, especially research that would extract several properties that have been recognized as essential in the eyes of the user, present these to information professionals working with digital objects, and compile concrete ideas of how these professionals would begin to digitally preserve these properties. Having a sound understanding of the processes, workflow, and models involved in digital preservation is important, as is an understanding of fundamental archival principles. It is also important to be familiar with the previous research on the subject of significant properties of digital objects because responsibly preserving digital objects requires an understanding of what their significant properties are. Research in digital preservation has begun to address the importance of the preservation of significant properties of digital objects; however, it is clear that there are some gaps in the current body of research on digital preservation. It is my hope that working to answer my research questions—what specific steps can the information professionals involved in digital preservation at the university's digital institutional repository take to make sure that these properties are preserved for the long-term, and how is the university's digital institutional repository infrastructure fostering digital preservation—will start to fill some of this gap.

## **Method**

This study used semi-structured interviews with information professionals and researchers directly and indirectly involved in the development of a new digital institutional repository at a research-oriented public university to gather perspectives on digital preservation and similar issues relevant to the field of library and information science.

### **Participants**

Eight participants were interviewed for this study. Five of these are information professionals working in an academic or special collections library environment and carry job titles such as systems specialist, applications analyst, archivist, or librarian. The remaining three are researchers or academics in the field of information and library science and archival studies and carry job titles such as professor or research assistant. The participants were recruited because of their involvement with the university's digital institutional repository. A total of twelve information professionals and researchers working with the university's digital institutional repository were invited to participate; however, only eight participants were interviewed. Because of the participants' close professional ties with the repository, the library, and the university, there does exist the risk of deductive disclosure once the outcomes of the study, in the form of the researcher's Master's paper, has been published. All participants were fully informed of this risk and gave their consent to be interviewed.

### **Research Materials**

An interview guide (see appendix A) was used to collect data for this study. The guide consisted of thirteen questions and opened by inquiring about the participant's

professional background and duties in general. Then it moved on to specific questions on the participant's role within the context of the university's digital institutional repository, the technical infrastructure of the repository, digital preservation at the repository, and additional questions directly concerning the digital institutional repository and the participant's personal views and opinions on it.

### **Procedure**

The first step in this study was to contact the study population with an invitation to participate in the study. Twelve information professionals and researchers working on the university's digital institutional repository were sent an email (see appendix B) outlining the purpose of the study and the general study procedures that would take place, and inviting possible participants to contact the researcher for more information if they were interested in participating in the study. When a possible participant contacted the researcher for more information and to set up an appointment for an interview, the researcher responded with a more detailed explanation of what would take place during the interview, including the time frame for the interview (approximately one hour) and that the interview would be recorded if the participant provided his/her consent for the researcher to do so. At this time the participant was also asked to choose a place to hold the interview where he/she felt comfortable. If an appointment was arranged, the researcher sent the interview guide to the participant two or three days before the interview to provide him/her an opportunity to prepare.

During the interview, the researcher first gave the participant an information and consent form (see appendix C) detailing the study procedures and potential risks. The participant was asked to read through this form and then sign if he/she did in fact wish to

participate with the study. The participant was given a copy of this information form to keep.

Next the researcher read a short introduction to the participant, again informing him/her that the interview would be recorded and assuring them that the researcher would take steps to anonymize all data collected during the interview and delete identifiers that would be traceable back to the participant. The digital recording device was then set to record, and the researcher began asking the participant the questions from the interview guide.

During the interview, the researcher took notes on the participant's responses to the questions. Based on the semi-structured interview protocol, the researcher had the freedom to reword or reorder questions as deemed appropriate within the context of the study and interview. The researcher did ask additional questions or for clarification when the content of the interview prompted these or if there was a need for clarification on specific responses given by the participant. The interview was brought to a close by a short thank you by the researcher as well as a reminder that the participant could choose to withdraw from the study at any time, and if he/she did indeed wish to do so to please contact the researcher or her advisor.

After the interviews, the researcher listened to the recordings and partially transcribed the interviews. These transcriptions were anonymized and then printed out for the researcher to review and annotate. The researcher also created a spreadsheet that contained condensed responses to each question asked in the interview. The researcher used the transcriptions and the spreadsheet to analyze the data collected through the study. The results gathered from this analysis are found in the next section of this paper.



## Results

This research study used semi-structured interviews with participants both directly and indirectly involved in the development of a new digital institutional repository at a research-oriented public university to gather data on many of the issues surrounding institutional repositories. Eight participants were interviewed for this study. Five of them are considered information professionals working in an academic or special collections library environment and carry job titles such as systems specialist, applications analyst, archivist, or librarian. The remaining three are researchers or academics in the field of information and library science and archival studies and carry job titles such as professor or research assistant. Because of the differing backgrounds of the participants interviewed, the opinions, priorities, and views on digital institutional repositories and digital preservation were naturally diverse and interesting. This section of the paper will describe who the participants were in terms of their professional backgrounds and duties and then describe what they said regarding the specific digital institutional repository that they are involved in developing, its goals, and how they think these goals will be accomplished.

### **The Study Participants<sup>6</sup>**

Eight participants who are involved in the development of the university's digital institutional repository were interviewed for this study. Each participant's professional background naturally informs his or her views and opinions on the digital institutional repository and its purpose and provides a relevant context for the information that will be presented in this paper; therefore, their backgrounds will be briefly described in this section.

The five information professionals interviewed for this study all work directly with the digital institutional repository. Three of the participants work within the university's library systems department. Specific to the digital institutional repository project, they identified their roles as project manager, software engineer/repository programmer, and project technical lead. Two of the participants work within the university's special collections library. Specific to the digital institutional repository project they identified their roles as project manager and operations officer.

The three researchers or academics interviewed for this study do not work directly with the digital institutional repository but do serve on the institutional repository steering committee. These three participants work as professors and researchers in the field of library and information science, more specifically in archives and records management, digital preservation, and data management.

The table below presents the job titles and functions of all study participants for ease of reference and comparison.

---

Table 1

*Participant's job title and job functions specific to the institutional repository project*

Job Title	Institutional Repository Job Function
Systems Librarian	Project Manager
Systems Specialist	Project Technical Lead
Applications Analyst	Software Engineer/Programmer
Archivist	Project Manager
Librarian	Operations Officer

Professor	Research/Steering Committee
Professor	Research/Steering Committee
Research Assistant	Research/Steering Committee

## **The Digital Institutional Repository**

### **Preservation as purpose.**

Over the course of the interviews with the eight participants many views, opinions, and ideas on the university's digital institutional repository and its past, present, and future were expressed. There was however one overarching concept that came through very clearly in every interview, and this concept can be seen as the driving force behind the university's digital institutional repository and integral to its mission and purpose. Every participant discussed the idea of preservation and how the repository can and will preserve the digital assets or objects that will be maintained or archived by the digital institutional repository. Preservation was discussed to varying degrees of emphasis or detail by every participant in the study.

When asked the question "What would you say is the main purpose of the digital institutional repository?", seven of the eight participants answered the preservation of, or preservation of and access to, the digital assets of the university. The remaining participant did not refer to preservation when asked this specific question but did further along into the interview. One response to this question was extremely telling, as the participant simply stated, "preservation, in a word" (Participant A, interview). Other statements such as "The main purpose of the [digital institutional repository] right now is to preserve objects that do not have a trustworthy home at the moment...", "I think the

main purpose is the preservation and access, so preservation of and access to, the significant content produced by university faculty, staff, and students”, and the “preservation of digital content for the [university] community” further support the claim that the most compelling reason and motivation for developing a digital institutional repository at the university where the study was conducted is to ensure the preservation of digital objects (Participant B, interview; Participant G, interview; Participant D, interview). These and other comments made by all the participants make it clear that the development effort has emphasized and encouraged preservation as the premise for the digital institutional repository and that they have researched and instituted a technical infrastructure that is intended to support the preservation of the digital assets of the university.

***What will be preserved?***

***Bit for bit.***

Going into more detail on the theme of preservation, all of the participants categorized as information professionals explained that by preservation they mean that the digital institutional repository is fundamentally “concerned with preserving the bitstream above all else” or “giving back the bits” (Participant A, interview; Participant D, interview). They all expressed that the primary focus of the digital institutional repository will be to “faithfully represent what was ingested,” and that further presentation and access applications for the repository are not currently being highlighted or prioritized within the development effort (Participant D, interview).

*Significant Properties.*

None of the participants, regardless of their professional background or research interests, felt comfortable more concretely defining what they would identify as the most significant properties of the digital objects that the digital institutional repository will preserve. However, most of the participants commented on the importance of asking this question and prompting the university's community and future depositors to the digital institutional repository to think about this topic. When asked to provide specific examples of significant properties of the digital objects that the digital institutional repository will manage, one participant cited "content, context, and structure" as the essential elements that will need to be preserved, and another participant cited "provenance, changes that an object has gone through over time as well as where it originated and any sort of associated context" (Participant D, interview; Participant H, interview). It was also suggested that

This is a question for the people contributing the content from around campus...I think the creator of the content, if asked in a non-technical way, should be the most qualified to think about what they would like to see preserved...I don't think there are generally one set of characteristics [that could be defined as significant properties]." (Participant G, interview).

So while the participants did not give specific examples of the significant properties of digital objects that will need to be preserved by the university's digital institutional repository, they did underscore what has already been articulated through the literature: that the concept of significant properties is a very complicated concept to define and to address. The participants instead gave general areas to consider or look within to answer this question more completely for digital objects that will be ingested into and then preserved by the repository.

*At-risk objects and collections.*

Most of the participants interviewed for this study indicated that the university's digital institutional repository needs to preserve at-risk or vulnerable digital objects. Although the research, planning, and development phases of the university's digital institutional repository has spanned at least five years, all of the participants referred to the urgency of the project because of the need to have a trustworthy repository where digital objects that have been designated as at-risk can be managed and preserved. This includes objects and existing digital collections that are susceptible to loss, damage, or obsolescence, objects and collections that are "hard to replace, hard to re-digitize," and even objects and collections that are currently managed by an individual university department or professor that does not "have the means" or acumen to keep them safe (Participant B, interview; Participant C, interview). One participant also noted the digital institutional repository as a place to "buy time for objects that cannot currently be preserved or...objects that need additional massaging before they can be [ingested and preserved by the repository]" (Participant F, interview). As evidenced by these statements, the claim for the need for the digital institutional repository to preserve at-risk or vulnerable digital objects and collections is strong. This tone of urgency is also reflected in the current digital preservation literature in which there is much discussion of format obsolescence, the imminent danger of loss or damage to digital media and objects, and the threat that if actions are not taken immediately there will be a great loss of digital content and a resulting gap in documentary materials from the current era.

*Moving forward and beyond.*

Another category of resources that some of the participants interviewed for this study classified as needing to be preserved by the university's digital institutional repository are digital objects and collections that are not currently being preserved or managed by the traditional library or even traditional digital repository. This includes scientific and social scientific data collections. Many participants expressed optimistic ideas and a desire to "move beyond" and create a truly innovative repository (Participant F, interview). These same participants also cited technological limitations and a need for resources, both of which are currently preventing the development of the university's digital institutional repository.

Although none of the participants interviewed for this study presented limitations as to the types of digital objects and collections or digital media formats that will be preserved in the university's digital institutional repository, it is clear that current technological boundaries and limited resources have presented themselves as obstacles in the development of the repository. All of the areas discussed by the participants—bitstream preservation, questioning and exploring the significant properties of digital objects and further ways to identify these, a recognized obligation to preserve at-risk digital objects and collections, and a commitment to innovation—convey that thoughtful planning and development practices have been employed thus far. Additionally, strategic positioning, an extensible framework, overall flexibility, and a spirit of innovation have been used to construct a strong platform for building a preservation-centric digital institutional repository. These factors will be discussed further in the next section of this paper.

***How it will be preserved.****The technical infrastructure.*

Further supporting the university's digital institutional repository as a preservation repository is the technical infrastructure that has been put in to place. This infrastructure was described by the participants involved in its development as using replication, fixity checks, and audit trails to preserve the digital objects and collections managed by the repository. All of the digital objects within the repository will be replicated across a distributed system; checksums will ensure bitstream preservation and the fixity of digital objects over time; and audit trails and mechanisms for generating, preserving, and tracking events and actions carried out upon the digital objects will be implemented. Some of the participants also mentioned an interest in further exploring the feasibility and necessity of format migrations, although one participant stated that the repository personnel "haven't really defined what that means" (Participant B, interview). The architecture of the university's digital institutional repository is clearly rooted in preserving the digital objects that the repository will manage, and an implementation that uses replication, fixity checks, and audit trails supports a preservation repository.

*Policy development.*

Additional ways to ensure that the digital objects that are being managed by the university's digital institutional repository will be preserved include establishing whether or not they are even worth preserving, and making decisions about whether the repository can guarantee that the object will be preserved. It is not feasible for objects to be appraised at the item level, so collecting policies will need to be established before the repository goes into full production. Several of the participants stated a need for such



policies to be established and a need to take “draft policies and make firm policies” (Participant G, interview). Other policies will need to be instituted to address the management of the digital objects themselves. Some examples of needed policies that were suggested by participants were those that address the technical needs of the digital institutional repository such as procedures for monitoring the health of the repository and core documents that define minimum ingest requirements (Participant B, interview; Participant A, interview; Participant F, interview). Defining clear policies will be important in maintaining and communicating that the university’s digital institutional repository can and will carry out its mission of preserving and maintaining access to the digital assets of the university.

*Creating a sustainable model.*

Throughout the course of interviewing the participants for this study, it also became clear that in order to fully commit to and carry out the preservation of digital objects, the repository will need to create a sustainable model for managing its tangible, intangible, and human resources. Maintenance of the technological infrastructure; training, engaging, and promoting the repository; and recruiting and retaining staff that is committed to its mission all require resources from library and university administration.

According to most of the participants interviewed for this study, the most successful and proven path to a sustainable future for the digital institutional repository will be to effectively communicate its value and worth to the university community. However, several of the participants implied, and one of the participants said, that the repository is “still getting there... We don’t have a sustainable organizational structure or governance yet [and] that’s a big missing component” (Participant B, interview).

Another participant stated that the repository must “bring in new allies and launch additional partnerships” working to remain “engaged [and] visible” (Participant F; interview). This same participant also stated that once people are using it and finding value in it, it will become an essential part of the university and part of the “bloodstream of the library” (Participant F, interview). Another participant expressed the same idea, stating that “the overall sustainability [of the repository] will come with use and access” (Participant B, interview).

Creating a sustainable model for the university’s digital institutional repository is integral to ensuring a trustworthy repository capable of preserving digital objects over the long term and is also one of the biggest challenges that the repository currently faces. The participants interviewed for this study have acknowledged this and have presented many ideas for paths to creating a more sustainable future for the repository.

*The repository as a research project.*

One last aspect of the university’s digital institutional repository effort that many of the study participants view as key in its ability to establish itself as a trustworthy and innovative preservation repository is that it is seen as “both a research project and a production system” (Participant A, interview). Most of the study participants commented on the extensive research and development that has gone into the repository thus far, and reflected that this is what will make the repository so powerful and innovative in opening opportunities for growth, collaborations, and the chances to become “a leader” in the field of digital preservation and digital and institutional repositories (Participant G, interview). Creating an environment where innovation and research are encouraged will greatly help to develop and foster the infrastructure and management of the university’s

digital institutional repository; it will also ensure the commitment of resources to, and longevity, sustainability, and growth of the repository.

### **Discussion**

The findings presented above clearly show that the university's digital institutional repository has been designed and will be implemented as a preservation repository for the digital assets of the university. Most of the information professionals and steering committee members working with the repository have identified the goal of the repository to be preservation and are taking steps to ensure that this goal is fully addressed and met. Decisions have been made about what will be preserved: the repository will specifically focus on preserving the bitstream of the digital object exactly as it was ingested. Additionally, the technical infrastructure of the repository has been designed to support preservation. Replication, fixity checks, and audit trails will help accomplish this goal.

The information professionals and steering committee members working with the university's digital institutional repository have also identified the need to further define and develop policies that will support the goals of the repository. These include policies addressing the issues of collection development and the appraisal of the digital objects that the repository can and will manage and policies addressing the technical needs of the repository.

The need to create a sustainable model for managing the tangible, intangible, and human resources of the digital institutional repository was also indicated as an important step in establishing and communicating the trustworthiness of the repository and its ability to preserve digital objects. Although steps are being taken in this direction, the

information professionals and steering committee members indicated this as one of the biggest challenges currently facing the repository and its staff.

Together with the small sample size of eight participants, conducting research within the context of a single digital institutional repository were the major limitations of this study. The condensed timeframe in which the research was conducted also limited the types of data that could be collected and prevented the researcher from being able to take a broader look at the development, past and present, of the digital institutional repository and to show expanded or comparative results.

These limitations do present opportunities for future research on digital institutional repositories in which it would be useful to study more than one repository over a longer timeframe in order to form a comparative view of these repositories. This would allow the researcher to evaluate possible similarities and differences in the repository's goals, and how these goals are being accomplished. A longitudinal study of a single digital institutional repository from initial planning to launch would also be interesting, as would a revisiting of the repository already studied to see how things panned out over time. Other opportunities for future research lie in evaluating how the users of digital institutional repositories access and use digital objects, or even how the creators of digital objects feel about preserving these objects within a repository.

This study set about to answer questions on how information professionals can ensure the preservation of digital objects within a digital institutional repository. Specific steps to ensure the long-term preservation of digital objects and how a digital institutional repository's technical infrastructure can be developed to foster digital preservation were conveyed. Perhaps most interesting, however, was how human dynamics, institutional

policies, and the creation of a sustainable repository model are often the most important factors to ensuring the preservation of digital objects over the long-term.

### Notes

1. JISC is an “independent advisory body that works with further and higher education by providing strategic guidance, advice, and opportunities use to use information and communication technologies to support learning, teaching, research, and administration” (<http://www.jisc.ac.uk/>).
2. Cedars stands for CURL Exemplars in Digital Archives.
3. An OAIS or ISO: 14721:2003 repository is “an archive consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community” (Consultive Committee for Space Data Systems, 2002, p. 1-1).
4. <http://www.significantproperties.org.uk>
5. This article reports the methodology and initial results of a multiple part study. While the authors cite the summer of 2009 as the scheduled publish date of the final results, I could not yet find these.
6. For the purpose of anonymity the participants will be referred using single-letter identifiers. These do not correspond in any way to the actual initials of the participants or the order in which they were interviewed.

## References

- About InSPECT*. (2007). Retrieved from  
<http://www.significantproperties.org.uk/about.html>
- Abrams, S. L. (2005). Establishing a global digital format registry. *Library Trends*, 54 (1), 125-143.
- Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., & Tsakonas, G. (Eds.).  
*Conference*. Berlin: Springer.
- Berninger, P., Brady, F., Hofmann, H., & Schram, J. (Eds.). (2002). Proceedings of the  
 DLM- forum 2002: @Ccess and preservation of electronic information: Best  
 practices and solutions. Luxembourg: Office for Official Publications of the  
 European Communities.
- Beyer, H. and Holtzblatt K. (1998). *Contextual design: Defining customer-centered  
 systems*. San Francisco: Morgan Kaufmann Publishers.
- The Cedars Project Team. (2001). *The Cedars project report*. Retrieved from  
<http://www.webarchive.org.uk/wayback/archive/20050410120000/http://www.leeds.ac.uk/cedars/admin/CedarsProjectReportToMar01.pdf>
- Consultive Committee for Space Data Systems. (2002). *Reference model for an open  
 archival information system (OAIS)*. Washington, D.C.: National Aeronautics and  
 Space Administration

- Dorner, D.G. (2009). Public sector readiness for digital preservation in New Zealand: The rate of adoption of an innovation in records management practices. *Government Information Quarterly* 26 (2009), 341-348.
- Holdsworth, D. & Sergeant, D. M. (2000). *A blueprint for representation information in the OAIS model*. Retrieved from <http://www.personal.leeds.ac.uk/~ecldh/cedars/nasa2000/nasa2000.html>.
- Hedstrom, M. & Lampe, C. (2001). Emulation vs. migration: Do users care? *RLG DigiNews* 5 (6). Retrieved from <http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2448.html#feature1>
- Hedstrom, M. L., Lee, C. A., Olson, J. S, & Lampe, C. A. (2006). “The old version flickers more”: Digital preservation from the user’s perspective. *The American Archivist*, 69 (Spring/ Summer), 159-187.
- Joint Information Systems Committee. (2006). *Digital preservation: Continued access to authentic digital assets*. Retrieved from <http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf>
- Knight, G. (2008). *Framework for the definition of significant properties: InSPECT work package 3.3*. Retrieved from <http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf>



- National Archives of Australia. (2002). *National archives green paper: An approach to the preservation of digital records*. Canberra: National Archives of Australia.
- Snow, K., Ballaux, B., Christensen-Dalsgaard, B., Hofman, H., Hofman Hansen, J., Innocenti, P., ... Thøgersen, J. (2008). Considering the user perspective: Research into usage and communication of digital information. *D-Lib Magazine*, 14 (5/6). Retrieved from <http://www.dlib.org/dlib/may08/ross/05ross.html>
- Thomas, S. & Martin, J. (2006). Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives. *Journal of the Society of Archivists* 27 (1), 29-56.
- Wilson, A. (2007). *Significant properties report: InSPECT work package 2.2*. Retrieved from [http://www.significantproperties.org.uk/documents/wp22\\_significant\\_properties.pdf](http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf)

## Appendix A

### Interview Guide

1. Could you tell me about your job, in terms of what do and your main responsibilities?
2. What is your specific role within the [INSTITUTIONAL-REPOSITORY-NAME] development effort?
3. What sorts of expertise would you say you bring to this process?
4. What would you say is the main purpose of [INSTITUTIONAL-REPOSITORY-NAME]?
5. What sorts of digital objects do you think the [INSTITUTIONAL-REPOSITORY-NAME] will manage?
6. Can you give me some specific examples?
7. Could you provide some examples of significant properties of these digital objects? In other words, what aspects of these objects do you feel are the most important to preserve? Think about quality, usability, rendering, and behavior.
8. How do you think the university's digital institutional repository will preserve the properties that you just highlighted?
9. Not everything can be preserved. How would you hope that the [INSTITUTIONAL-REPOSITORY-NAME] will address this issue?
10. What do you see as the main priorities for the [INSTITUTIONAL-REPOSITORY-NAME] team over the next one to two years?

11. What do you envision as the biggest challenges for the [INSTITUTIONAL-REPOSITORY-NAME] in ensuring long-term access to the materials that it ingests?
12. How do you think the [INSTITUTIONAL-REPOSITORY-NAME] team will address those challenges? Please be specific.
13. Is there anything else I should have asked you about [INSTITUTIONAL-REPOSITORY-NAME]? Or is there anything else that you'd like to tell me?

## **Appendix B**

### **Invitation to Participate**

Dear Subject,

I am a M.S.L.S. student at the School of Information and Library Science, and I am writing my master's paper on digital preservation at the [INSTITUTIONAL-REPOSITORY-NAME]. My research study will consist of meeting with and interviewing information professionals at the [INSTITUTIONAL-REPOSITORY-NAME] to get their thoughts on digital preservation and the associated policies and procedures that are being introduced at the [INSTITUTIONAL-REPOSITORY-NAME]. I was given your name and contact information as someone who is playing a significant role in the development of the university's [INSTITUTIONAL-REPOSITORY-NAME]. If you are indeed interested I would be happy to send you more detailed information on the study procedures. Please let me know.

Sincerely,  
Meg Tuomala

## Appendix C

### Consent Form

#### Information about a Research Study

---

Title of Study: Understanding Digital Preservation at the [INSTITUTIONAL-REPOSITORY-NAME]

Principal Investigator: Meg Tuomala, Graduate Student  
UNC-Chapel Hill Department: School of Information and Library Science  
Faculty Advisor: Cal Lee, Assistant Professor

Study Contact telephone number: 919.259.9643  
Study Contact email: mtuomala@gmail.com  
Faculty Advisor Contact telephone number: 919.962.7024  
Faculty Advisor email: callee@ils.unc.edu

---

#### What are some general things you should know about research studies?

You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

#### What is the purpose of this study?

To understand digital preservation at an institutional level by studying procedures locally at the [INSTITUTIONAL-REPOSITORY-NAME].

#### How many people will take part in this study?

If you decide to be in this study, you will be one of approximately 5-10 people in this research study.

#### How long will your part in this study last?

The interview will take less than one hour. You can choose to stop the interview at any time.

What will happen if you take part in the study

I will ask you questions about your work at the [INSTITUTIONAL-REPOSITORY-NAME]. I will take notes about what you say and you will be recorded. You do not have to answer any questions that you do not wish to answer, for any reason. If you do not wish to be recorded let me know and I will take only written notes.

What are the possible benefits from being in this study?

Research is designed to benefit society by gaining new knowledge. Your participation is important to help us understand digital preservation at the institutional level and you may even benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study?

The only risk to you might be if your identity were ever revealed through deductive disclosure. This could result in professional embarrassment or even the loss of employment. To minimize this risk I will keep all identifiers separate from the data I am collecting today. There are no other expected risks to you for helping me with this study, but you and even the library can expect to benefit from this study as it could further inform your understanding of digital preservation.

How will your privacy be protected?

All the information I receive from you during this interview, including your name and any other identifying information, will be strictly confidential and will be kept under lock and key. I will not identify you or use any information that would make it possible for anyone to identify you in any presentation or written reports about this study. If it is okay with you, I might want to use direct quotes from you, but these would only be quoted as coming from “a participant” or “an information professional.” I will not associate any gender, age, profession, specific title, or other specific information with these direct quotes. I will not include the name of the institution or project in any report or presentation. When I finish with all the interviews from everyone who has agreed to participate, I will group all the answers together in any report or presentation. There will be no way to identify individual participants.

Will you receive anything for being in this study?

I am not going to pay you for your information, but your information is very important to us.

Will it cost you anything to be in this study?

There are no costs for being in the study other than your time.

What if you have questions about this study?

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact me at mtuomala@gmail.com or at 919.259.9643.

What if you have questions about your rights as a research participant?

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu).

*Thank you for helping me with this study.*

-----

Title of Study: Understanding Digital Preservation at the [INSTITUTIONAL-REPOSITORY-NAME]

Principal Investigator: Meg Tuomala

Participant's Agreement:

I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

\_\_\_\_\_  
Signature of Research Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Printed Name of Research Participant

\_\_\_\_\_  
Signature of Research Team Member Obtaining Consent

\_\_\_\_\_  
Date

\_\_\_\_\_  
Printed Name of Research Team Member Obtaining Consent