The University of North Carolina at Chapel Hill (UNC-Chapel Hill) is exploring opportunities to support the teaching and research needs of its administrators, faculty, and students through the preservation and dissemination of its digital assets. Recent studies suggest that much of this intellectual output is at risk and may disappear unless stewards of this information emerge and take active steps to ensure its preservation for current and future use. This paper presents the results of a survey of 200 UNC faculty members that documents current processes and practices related to research materials. It serves as an important first step in a process designed to describe the needs and expectations of information producers on campus related to the collection, storage, and preservation of research materials.

Headings:

- Digital Preservation
- Collaboration
- Research
- Institutional Repositories
DATA WORTH PRESERVING: AN EXPLORATORY STUDY OF THE COLLECTION, STORAGE, USE, AND PRESERVATION OF RESEARCH DATA AT CAROLINA

by
John C. Schaefer

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
November 2006

Approved by

_______________________________________
Helen R. Tibbo
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Methodology</td>
<td>7</td>
</tr>
<tr>
<td>Findings</td>
<td>17</td>
</tr>
<tr>
<td>Discussion</td>
<td>33</td>
</tr>
<tr>
<td>Conclusion</td>
<td>42</td>
</tr>
<tr>
<td>References</td>
<td>44</td>
</tr>
<tr>
<td>Appendix A</td>
<td>52</td>
</tr>
<tr>
<td>Appendix B</td>
<td>57</td>
</tr>
</tbody>
</table>
Introduction

Currently, the University of North Carolina at Chapel Hill (UNC-Chapel Hill) is exploring opportunities to support the teaching and research needs of its administrators, faculty, and students through the preservation and dissemination of its digital assets. A growing number of universities view digital content as an institutional asset that should be managed and preserved (Hey, 2004; Pritchard, Carver, & Anand, 2004; Lynch, 2003; Rogers, 2003). Recent studies suggest that much of this intellectual output is at risk and may disappear unless stewards of this information emerge and take active steps to ensure its preservation for current and future use (Hedstrom, 2003; Hedstrom & Ross, 2003). Therefore, like other leading academic research institutions (e.g., Massachusetts Institute of Technology (MIT), University of Michigan, University of California Santa Barbara, Ohio State University), the university seeks to implement new technology and develop new processes, policies, and procedures to manage these valuable assets effectively and preserve them for future use.

To facilitate this process, the university created an interdisciplinary task force of subject matter experts and key stakeholders to investigate both digital curation and institutional repositories. This effort is now focused on planning and implementing a pilot institutional repository and preservation program. The committee will use the pilot’s results to facilitate the development of policies, procedures, and long-term strategies to prepare for a future campus-wide implementation (Tibbo, 2005, March).
A key step in this effort is to determine the current processes and practices of UNC-Chapel Hill faculty members as they relate to the collection, use, storage, and preservation of research data. The wealth of material that exists documenting the experiences of organizations exploring the use of institutional repositories suggests that developing an understanding of the current practices, needs, and expectations of information producers is a critical first step in this process (Barton & Waters, 2005; Baudoin & Branschofsky, 2004; Branin, 2004; Genoni, 2004; Rogers, 2003).

This paper presents the results of a survey of 200 UNC faculty members designed to document their current processes and practices related to research materials. It is intended to serve as an input to the first step in describing the needs and expectations of information producers on campus related to the collection, storage, and preservation of research materials. Specifically, this exploratory study seeks to answer the following questions:

- What types of data do UNC-Chapel Hill faculty members collect during the research process and what formats do they store it in?
- To what extent do these researchers collaborate on projects? Where are their collaborators located and how do they share data?
- What are the research data preservation needs of Carolina researchers? How are these needs currently met by the university, schools, or departments?

**Background**

**Research in University Settings**

Today’s universities are centers of learning as well as institutions that create solutions to real-world problems. In the United States, the federal government invests
heavily in sustaining and expanding the research capacity of these institutions, providing funds to support the research efforts of academics in a broad range of disciplines, including the sciences, social sciences, and the humanities. In turn, the universities rely on the ability of their academics to attract funds to sustain and grow their institutions. Most universities consider an academic’s ability to win competitive grants and his or her participation in funded research projects as evidence of scholarly activity, which plays a role in tenure-related decisions (Blackburn & Lawrence, 1995; Clark, 1995; Fechter, 1999).

**Impact of Information Technology on the Research Process**

The growing use of information technology (IT) in the academic research process is profoundly changing research practices. The new technology facilitates collaboration and interdisciplinary research, supports the capture, use, and reuse of data, enables new methods to analyze data, and increases access to this information. In particular, the use of IT provides a richer, more productive research environment by enabling scholars to utilize new tools to facilitate the exchange of data (Houghton, Steele, & Henty, 2004; Brockman, Neumann, Palmer, & Tidline, 2001).

In the sciences, technological advances have in some cases reduced the need for traditional laboratory experiments monitored by white-coated lab assistants armed with clip boards and paper-based check-lists. Instead, scientists now conduct virtual experiments using advanced modeling, simulation, and visualization techniques. These efforts result in large amounts of data that researchers must manage throughout a project’s lifecycle and in some cases preserve far into the future to meet specific
conditions set by funding agencies (Hodge & Frangakis, 2004; Jacobs & Humphrey, 2004).

The ubiquitous nature of technology in the Digital Age has also impacted researchers in the fine arts, humanities, and social sciences. In fact, Waters (2003, December) described scholars in the humanities as “video, image, and sound hungry” (9). Like their counterparts in the sciences, scholars in these areas benefit from advances in text encoding, multimedia, and remote sensing technologies (Brockman, Neumann, Palmer, & Tidline, 2001). Archaeologists, for example, collect a wide variety of materials in various forms to facilitate their studies. Traditionally a document-intensive field of study that relied heavily on analog materials, archaeologists now use digital technologies to capture accurate data in the field, including geographic positioning systems (GPS), databases, and digital photography (Waters, 2003, December).

As a result of the incorporation of IT advances into the research process, universities are currently awash in enormous amounts of research data. In the case of research materials in digital format, recent studies indicate that much of this valuable material is at risk of disappearing forever (Hedstrom, 2003; Hedstrom & Ross, 2003).

**Role of IRs in Capturing and Disseminating Research**

Institutional repositories (IRs) provide universities with an effective means to capture, preserve, and provide access to the intellectual output of their scholarly communities (Branin, 2004). These repositories consist of both the underlying information technology and services provided by the institution to support the needs of its community (Crow, 2002; Lynch, 2003).
In addition to the technical infrastructure, they consist of policies, procedures, and negotiated agreements that define specific services and set performance indicators. Developing this documentation requires close collaboration between producers, data managers, administrators, and end users (Baudoin & Branschofsky, 2003; Lynch, 2003).

The development of most information systems typically follows a specific methodology to ensure success. A key component of the methodology consists of developing a detailed understanding of the current environment. In a research environment, this entails working closely with the creators or producers of the content that will populate the new system. The interaction is crucial to ensure that the new system will meet the needs and expectations of the targeted users.

In the case of institutional repositories, an understanding of the current environment of the designated community of producers is critical step in the development and implementation process. Barton and Waters (2004) highlight the importance of interacting with producers to determine their needs in order to define institutional repository services, including the types of content that the repository will support, preserve, and provide access to over the long-term. In addition, a leading standard for institutional repositories, the Open Archival Information System (OAIS) reference model, clearly states the need for close collaboration between information producers and managers (Consultative Committee for Space Data Systems, 2002; Lavoie, 2004). Through this interaction, the information producers gain the opportunity to clearly express their needs and expectations.
Methodology

This exploratory study consisted of two phases. In phase one, I tested a participant selection process and solicited feedback from 25 UNC-Chapel Hill faculty members on a survey designed to document practices related to the storage of research data. This phase served as a pilot study and provided detailed information that contributed to changes to both the participant selection process and the survey instrument. Phase two consisted of the implementation of an institution-wide study of 200 faculty members randomly selected from the university’s schools and divisions.

To provide the context for the study’s design, this section first describes its population, including the structure used by the university to organize its academics and researchers. This structure facilitated the selection of study participants. Next, it reviews the data collection strategy employed in this study, which enabled me to solicit and capture relevant information from UNC faculty members from a broad range of disciplines. The section then discusses the pilot project before closing with a detailed description of the university-wide effort.

Study Population

UNC-Chapel Hill possesses a rich tradition of scholarly research. Its academics, scientists, researchers, and students routinely advance the field of knowledge in their disciplines. The university’s strong support for and encouragement of the research process has resulted in national and international recognition of its faculty members for
their contributions to the advancement of knowledge in their disciplines (http://research.unc.edu/resfacts/faculty.php). As shown in Table 1, the university organizes its 2,871 faculty members into two divisions: Academic Affairs and Health Affairs. This structure illustrates the parameters of the sample frame for this research effort.

Table 1

Organization of UNC Chapel Hill faculty members by division

<table>
<thead>
<tr>
<th>Academic Affairs</th>
<th>Health Affairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ College of Arts &amp; Sciences (<em>Fine Arts and Humanities; Sciences; Social Sciences</em>)</td>
<td>□ School of Dentistry</td>
</tr>
<tr>
<td>□ Kenan-Flagler School of Business</td>
<td>□ School of Medicine</td>
</tr>
<tr>
<td>□ School of Education</td>
<td>□ School of Nursing</td>
</tr>
<tr>
<td>□ School of Government</td>
<td>□ School of Pharmacy</td>
</tr>
<tr>
<td>□ School of Information &amp; Library Science</td>
<td>□ School of Public Health</td>
</tr>
<tr>
<td>□ School of Journalism &amp; Mass Communication</td>
<td></td>
</tr>
<tr>
<td>□ School of Law</td>
<td></td>
</tr>
<tr>
<td>□ School of Social Work</td>
<td></td>
</tr>
<tr>
<td>*Total Faculty: 1,164</td>
<td>*Total Faculty: 1,707</td>
</tr>
</tbody>
</table>

*Total Faculty consists of the following categories: tenured, tenured-track, and fixed termed. Source: Office of Institutional Research and Assessment, Fact Book, 2005-2006, Table 27 Full-time Permanent Faculty, 2005 (http://www.ais.unc.edu/ir/factbook/fb2005-2006/faculty_staff/fb05tbl27.html)

Guided by this sample frame, I selected 25 full-time UNC faculty members for the pilot project. The university-wide study involved 200 full-time faculty members, with 100 individuals picked from each division to provide broad coverage of the various research practices used on campus. Faculty members who decided to participate in the study were asked to answer questions based on their research experiences over the last
three years. I selected this timeframe to take into account the impact of information technology advances on the research process.

Data Collection Strategy

Survey Research

Surveys provide researchers with an economical method to collect standardized data from a large, dispersed population in a short period of time. In particular, survey research affords researchers with an ideal tool to capture data necessary to describe characteristics, habits, or practices of a population. The use of surveys, however, limits researchers’ direct contact with participants and restricts their ability to ask follow up questions or introduce additional questions to pursue new areas of interests. Furthermore, surveys must be brief and simple to encourage completion thus limiting the quantity and quality of responses. Finally, researchers must have an in depth understanding of their audiences to design a useful tool (Babbie, 2004; Covey, 2002).

For the purposes of this study, survey research provided an effective and efficient method to collect data necessary to determine how UNC-Chapel Hill faculty members handle research data. The strengths of this strategy matched well with the study’s objectives. For example, it facilitated the collection of standardized data from faculty members in different disciplines. As demonstrated in the findings section, this data enabled me to develop an informative picture of their practices in this area and supported comparative analysis. Finally, the minimal contact and effort required to complete the brief online survey more than likely encouraged individuals to participate in this study. While the limited contact with participants precluded the opportunity to ask follow-up questions, the high-level of participation and resulting data provided me with the ability
to determine the practices and habits of a large number of UNC faculty members related to the collection and storage of research data.

**Survey Instrument**

*Description.* The survey instrument (see Appendix A) consisted of four sections: 1) collection and storage of data, including six multiple choice questions about materials that faculty collect, manage, and store during the research process; 2) collaboration and sharing of data, including one yes or no question and four multiple choice questions; 3) data preservation needs, including one multiple choice question and two open-ended questions; 4) demographics, consisting of one question asking survey participants to select their primary school or college from a pick-list.

*Structure.* The survey’s structure was based on a set of questions used to explore the research practices of University of California, Santa Barbara (UCSB) faculty members and on a review of institutional repository websites and related literature. In 2004, Pritchard, Carver, and Anand interviewed UCSB faculty members to discover the characteristics and needs of data-intensive research projects in the sciences. They expanded their effort to include researchers from other fields such as education, humanities, and fine arts as a control group. In particular, the UCSB study collected data regarding data collection and storage, sharing, and long-term preservation (Pritchard, Carver, & Anand, 2004). A review of this information facilitated the development of similar categories for this effort.

The review of institutional repository websites and related literature provided useful information about the formats faculty members use to store research data. Specifically, format support policies maintained by The Massachusetts Institute of
Technology (MIT) (http://libraries.mit.edu/dspace-mit/build/policies/format.html) and the University of Michigan (http://deepblue.lib.umich.edu/about/deepbluepreservation.jsp) provided useful information about the types of formats used by their faculty members. Information from these institutions contributed to the development of pick-lists and examples to include in the survey (see Table 2).

Table 2

<table>
<thead>
<tr>
<th>Formats included as examples in the survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text (e.g., HTML, Text, Rich Text Format, XML, SGML, ASCII)</td>
</tr>
<tr>
<td>Image (e.g., GIF, JPEG, PNG, TIFF, BMP, Photo CD)</td>
</tr>
<tr>
<td>Audio (e.g., AIFF, MPEG, RealAudio, WAV)</td>
</tr>
<tr>
<td>Adobe PDF</td>
</tr>
</tbody>
</table>

Interviews

The use of an interview strategy presents researchers with an effective means to engage study participants. When conducted in person, interviews afford researchers with an opportunity to clarify questions and probe responses (Babbie, 2004; Yin, 1984). In the case of this study, the interviews took place in the participants’ work environment, which provided the additional opportunity to view their work space and, if present, physical storage locations such as file cabinets, bookshelves, or piles of paper on tables.

This direct interaction played a crucial role in the pilot study (i.e., phase one). As described in more detail in the next section, the faculty members who took part in this effort first completed a survey and then participated in a personal interview to discuss the results. In these focused interviews, participants provided their answers to survey questions and offered suggestions on how I could revise the questions to improve clarity or relevancy. They also suggested new questions to fill gaps in the survey instrument.
Although the large number of potential participants (i.e., 200) precluded the use of this strategy in the university-wide study (i.e., phase two), the data obtained from the interviews led to significant improvements in the survey instrument and ultimately in the quality and accuracy of the data collected.

**Pilot Study (Phase One)**

Project management principles and best practices suggest that pilot studies play an essential role in the successful completion of efforts in a variety of fields. These efforts provide managers, team members, and subject matter experts with the ability to test policies, procedures, technology, and other components critical to the success of a project in a controlled environment and at a reduced scale. When a project’s success is based on the collection of accurate, relevant data, pilot studies present principle investigators, project teams, independent researchers or others with an opportunity to test the quality and appropriateness of data collection tools prior to using them in a large-scale study. In the case of surveys, this is an essential step to ensure that the questions are precisely worded and capture data appropriate to the research effort (Covey, 2002).

Completed in the fall of 2005, the pilot study solicited feedback from 25 Academic Affairs and Allied Health faculty on the survey instrument. Specifically, the questions sought information about the types and formats of the data that they collect during the research process. It also included questions regarding digital storage and preservation needs as well as challenges faced by faculty in these areas. In addition, this effort afforded an opportunity to test the participant selection process and data collection efforts.
Selection of Participants

The selection of pilot participants entailed a two-step process. First, I reviewed the listing of academic schools and departments on the university’s website and randomly selected departments that I believed generated a large amount of research data. To ensure a broad representation of participants, I selected departments from a range of disciplines, including humanities, social sciences, and sciences. Next, I visited the specific website, navigated to the faculty directory, and reviewed the biographical data for each individual. In most cases, this information included research interests, recent publications, and short descriptions of ongoing efforts. I selected individuals that evidenced strong and current participation in the research process as potential pilot study participants.

Those individuals identified as potential participants received an email invitation to participate in the study including background information to place the study into context. A consent form and survey were attached to this message. Those individuals who elected to participate then completed the survey and were later debriefed to identify their issues, observations, and suggested improvements regarding the survey instrument and process.

Results of Pilot Study

Although the pilot ultimately involved interviews with just three participants (a 12% response rate), the rich information gleaned from the data provided by these individuals improved the process. For example, the results suggested that the initial focus of the survey on digital research data should be expanded to include non digital information. The debriefing sessions with the participants clearly suggested that researchers continue to rely on non digital materials to complete their research endeavors.
In a discussion related to storage needs, for example, one participant indicated the need for offsite storage to accommodate non digital materials, including paper and video items. Another participant mentioned that microfilm and other non digital materials continued to play an essential role in her research process. As a result of this feedback, I modified the scope of the survey to include questions related to non digital materials.

*University-wide Study (Phase Two)*

**Selection of Participants**

To select participants for phase two, I used a process similar to the one employed in the pilot phase with slight modifications to accommodate the larger sample size. First, I visited the university’s website and navigated to the listing of *Departments* ([http://www.unc.edu/depts/](http://www.unc.edu/depts/)). Using this web page as a portal, I then accessed the homepages of the schools or departments and systematically selected individuals from the faculty list using the same criteria tested during the pilot project. For each individual, I noted their name and email address in a spreadsheet to facilitate the distribution of invitations to participate in the study.

**Distribution of Survey**

Due to the size of the sample (i.e., 200 potential participants), I implemented the survey using an online service called *SurveyMonkey* ([http://www.surveymonkey.com](http://www.surveymonkey.com)). This service provides tools to facilitate survey development. It includes the ability to pick question formats (e.g., yes/no, multiple choice, and open-ended) and create logical paths through the survey based on previous answers. In addition, the online service provides features that protect the privacy and confidentiality of research subjects. For example, *SurveyMonkey* presents survey administrators with the ability to enable participants to
respond anonymously. Finally, the service includes flexible analytical, reporting, and export tools to facilitate data analysis.

The distribution of invitations to participate in this study was accomplished in three stages. The initial invitation (see Appendix B) was sent via email on Sunday, May 21, 2006 in two separate batches: the first batch consisted of 100 participants from Academic Affairs; the second batch consisted of 100 participants from Allied Health. Approximately 5% of the emails were returned as errors, which I corrected and sent as a subsequent message.

In an effort to increase the number of participants, I sent two reminder messages to the blind distribution list. This message consisted of the text of the original message (including the link to the survey) in addition to text requesting individuals to participate in this study. I sent the first reminder on Tuesday, May 30 (nine days after the original distribution) and the second reminder on Monday, June 5 (15 days after the original distribution). The survey was closed on Friday, June 9 (Day 20).

Analysis of Data

After the closing date, I exported the results into a Microsoft Excel Workbook. This step was necessary in order to manipulate the data at a lower level of granularity than supported by the survey tool. For example, while the date collection tool (i.e., SurveyMonkey) facilitated the creation of reports at a high-level, it did not support the generation of reports at a level of detail sufficient to support comparative analysis of the research practices of different groups of academics.

As previously noted, one of the objectives of this study is to determine if practices related to the collection, storage, and preservation of research data vary across
disciplines. Therefore, as illustrated in Table 2, I organized the study participants into two groups to facilitate comparative analysis of these practices. The findings presented in the next section reflect this organization structure.

Table 3
Organization of study participants to facilitate comparative analysis

<table>
<thead>
<tr>
<th>Fine Arts, Humanities, Social Sciences, and Professional Schools</th>
<th>Science and Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>- College of Arts &amp; Sciences (Fine Arts and Humanities; Social Sciences)</td>
<td>- College of Arts &amp; Sciences (Sciences)</td>
</tr>
<tr>
<td>- Kenan-Flagler School of Business</td>
<td>- School of Dentistry</td>
</tr>
<tr>
<td>- School of Education</td>
<td>- School of Medicine</td>
</tr>
<tr>
<td>- School of Government</td>
<td>- School of Nursing</td>
</tr>
<tr>
<td>- School of Information &amp; Library Science</td>
<td>- School of Pharmacy</td>
</tr>
<tr>
<td>- School of Journalism &amp; Mass Communication</td>
<td>- School of Public Health</td>
</tr>
<tr>
<td>- School of Law</td>
<td>-</td>
</tr>
<tr>
<td>- School of Social Work</td>
<td></td>
</tr>
</tbody>
</table>

# surveys sent: 90 # surveys sent: 110
# surveys completed: 23 # surveys completed: 35
% completed: 25 % completed: 38

Total Faculty consists of the following categories: tenured, tenured-track, and fixed termed. Source: Office of Institutional Research and Assessment, Fact Book, 2005-2006, Table 27 Full-time Permanent Faculty, 2005 (http://www.ais.unc.edu/ir/factbook/fb2005-2006/faculty_staff/fb05tbl27.html)
Findings

Participants

Of the 200 participants invited to participate in this study, 58 chose to complete a survey, resulting in a 29% response rate. Figures 1 and 2 below show the breakout of participants at the organizational unit level.
As shown in Figure 3, the majority of participants (33 or 56.9%) decided to take part in the study within two days of receiving the initial invitation. It is interesting to note that the highest rate of participation (18) occurred on a Sunday. Other than the first business day of the survey period, when 15 individuals completed the survey, the second and third highest levels of participation occurred on the days I sent reminder notices (i.e., Monday, June 5 (Day 15) and Friday, June 9 (Day 20)).

![Figure 3 Number of Surveys Completed by Day](image)

**Types of Data Collected During the Research Process**

To facilitate research, this study’s participants primarily collect data obtained through observations or measurements (86.2%). Slightly more than half of the researchers (53.4%) rely on documents or texts and a somewhat smaller percentage (43.1%) collect data in the form of images or films. Researchers also collect artistic or cultural artifacts (13.8%) and other types of materials (12.1%), including audio files, computer programs, datasets, patient information and follow-up data, and in-depth interview materials. One participant noted a difference in the use of the data, stating that
“pure research data is only observations and measurements” while “documents and texts [apply only] to the literature of other authors.”

**Figure 4 Types of Data Collected by Researchers (n=58)**

When analyzing the data at a lower level of granularity, the differences between the research processes of the two groups in this study become clearer. For example, the majority of researchers in sciences or health (97.1%) collect data in the form of observations or measurements, with images or films (37.1%) and documents or texts (28.6%) rounding out their data acquisition efforts. One participant also noted that he or she collects “computer code, often in MATLAB or C++.”

In contrast, humanists, social scientists, and researchers in the university’s professional schools appear to rely on a greater breadth of data types. While documents and texts formed the largest percentage (91.3%) of the types of data collected, many of these researchers also collected observations or measurements (69.6%), images or films (52.2%), and artistic or cultural artifacts (34.8%).
Formats Used to Store Research Data

The findings suggest that researchers store data in multiple formats. While most researchers (84.5%) commonly store data in text formats (e.g., text, rich text format, HTML, XML), they also note the use of other formats, including images (55.2%), databases (51.7%), and Adobe PDF (39.7%). The results provide an indication of the breadth and complexity of research conducted by UNC faculty members. For example, 17 researchers (29.3%) noted the storage of data in applications such as geographic information systems (GIS) or science/mathematic systems. Other data formats used to support the range of research conducted on campus include video (20.7%) and audio (10.3%).

Comments from several participants suggest that researchers select formats based on data type or intended use. For example, one researcher noted that he or she uses “proprietary formats for geophysics, stratigraphy and other data.” Another individual noted the use of highly specialized data acquisition and visualization software created by Plexon and National Instruments (i.e., LabVIEW). Finally, one participant qualified their
selection of formats by noting that “PDF-files only apply to literature; raw research data is stored as ASCII files … [and] after processing also as Excel or graph display programs.”

**Data Storage Locations**

**During Project Lifecycle**

*Number of locations used to store research data.* During its active life, the data suggests that UNC faculty members store their information in multiple locations. Generally, researchers in science and health store active research materials in 3.3 locations on average. In comparison, researchers in the fine arts, humanities, social sciences, or professional schools place their materials in four locations on average.

*Storage locations.* In addition to storing data on the hard-drives of UNC supported desktops (65.5%) or laptops (46.6%), a large number of researchers (65.5%) store data on department or school servers. Researchers (62.1%) also store active data on portable media (62.1%) and paper (41.4%).

Several participants provided additional information about their practices related to the storage of active data. The other storage locations include private ISP servers and sites maintained by public and private funding agencies. One response provides an example of the high level of technical competency possessed by many of the researchers in this study. In regard to the question, the individual noted that she runs a “server with daily backup of all computers… [in her] research group.”
When the results are investigated at a lower level, several differences between the practices of researchers in other disciplines become evident. For example, the data suggests that science and health researchers prefer to store active research data on their desktops (77.1\%) versus laptops (37.1\%). In comparison, humanists, social scientists, and researchers from the professional schools prefer the use of laptops (60.9\%) over desktops (47.8\%). These researchers also tend to store data on non UNC supported laptops at a higher percentage than science and health researchers (i.e., 26.1\% vs. 2.9\%). Finally, researchers from the humanities, social sciences, and professional schools tend to use AFS space at a higher rate than the other disciplines (i.e., 30.4\% vs. 5.7\%).

**After Project Completion**

*Why researchers maintain data past project completion.* As shown in Figure 7, the majority of researchers (84.5\%) save data for their own reuse. Over half (58.6\%) save data for potential reuse by others while less than half (44.8\%) save data due to requirements set by funding agencies, publishers, or the university. Few save data for other purposes such as public relations or outreach activities. Two respondents noted that
they save data to “verify or answer questions” or “just to have a record in case… [it is needed] to check quotations for publications, etc.”

**Figure 7 Reasons Why Researchers Maintain Data Past Project Completion (n=58)**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential reuse of data (personal)</td>
<td>80.0%</td>
</tr>
<tr>
<td>Potential reuse of data by others</td>
<td>60.0%</td>
</tr>
<tr>
<td>Required by funding agencies</td>
<td>40.0%</td>
</tr>
<tr>
<td>Public relations / outreach</td>
<td>20.0%</td>
</tr>
<tr>
<td>Other</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

*Materials required by funding agencies, publishers, or the university.* Researchers identified several types of materials that these entities require them to maintain past the completion of their endeavor. This information includes research materials (55.2%), articles and publications (27.6%), administrative records (25.9%), reports and presentations (20.7%), and project-related correspondence (13.8%). Several respondents took the opportunity to specify items, including “clinical records,” “database records,” and “signed IRB consent forms.” Interestingly, few respondents (3.4%) noted the need to store websites or portals beyond project completion.

*Number of locations used to store inactive data.* The data suggests that UNC faculty members store research materials in fewer locations after project completion. For example, science and health researchers indicated that they store materials from completed efforts in an average of 2.5 locations (versus 3.3 locations during active life)
and researchers in the humanities, social sciences, and professional schools noted the use of 3.3 locations (versus four locations during active life).

**Storage locations.** Generally, researchers continue to rely on hard drives of UNC-supported desktops, department / school servers, portable media, and paper as the primary locations to store research materials (see Figure 8). It is interesting to note that faculty in the sciences and health who participated in this survey show a slight increase in the use of paper to store materials from completed efforts.

![Figure 8 Data Storage Locations After Project Completion (n=58)](image)

**Collaboration During the Research Process**

The survey contained specific questions designed to assess the level of collaborative efforts taking place on campus as well as how these efforts impact the handling of research data. In regard to collaboration, the data clearly suggests that faculty routinely collaborate with their colleagues. When asked if they collaborate with other researchers, 51 of the 58 participants or 87.9% answered in the affirmative.

**Types of collaborations.** Most participants’ collaborative efforts consist of endeavors with colleagues within the same department (39.2%). Many of these efforts
cross departmental (19.6%) and institutional boundaries (23.5%). A smaller number of participants (6%) indicated that their collaborative research efforts within the last three years involved researchers from outside academia, to include public, private, and non profit institutions.

![Figure 9 Primary Locations of Research Collaborators](n=51)

**Methods used to share data.** Collaborative efforts require the sharing of data. As noted in recent studies, email technology greatly facilitates the ability of knowledge workers to share data and collaborate on projects (See for example Houghton, Steele, & Henty, 2004). Similarly, a majority of this study’s participants (74.5%) use email to share their research data. Other common methods of data sharing include person to person (51%), department/school server (45.1%), and portable media (39.2%). Even given the ubiquitous nature of information technology, a high number of participants (35.3%) continue to share data on paper to advance their research.

Additional comments shared by two respondents suggest that researchers might encounter some problems in regard to large file sizes or security issues. For example, a
participant noted that “if [I need to transfer] large files I usually upload them onto my AFS/ONYEN public html-folder and send an email [to my collaborators] with the link to the respective files.” When sharing sensitive data, a participant responded that he prefers to use a secure FTP but may resort to “hand delivery… or a reliable shipper” if required to protect the data.

Data storage after completion of collaborative efforts. Both groups of faculty members participating in this study store data from collaborative efforts primarily at their home institution (i.e., UNC-Chapel Hill). More than half of the researchers (51%) store data just at UNC while the others store data at both UNC and other sites (43.1%). Three participants (5.9%) provided personal responses to this question.

Comments from two researchers from the science and health communities suggested the use of servers maintained by entities other than their home departments. For example, one noted the he saves data from collaborative efforts “on a server for our hospital/AHEC” while the other said she stores data at the “Collaborative Studies Coordinating Center at UNC and George Washington University Biostatistics.”

In contrast, one researcher from the humanities, social science, and professional schools communities provided a response that might suggest the lack of specialized repositories in these disciplines. In response to the question, the research noted that he stores data from collaborative efforts on “my own media.”

Long-term Preservation Needs

Users’ Perspective of Current Needs

When prompted by an open-ended question about their long-term preservation needs, survey participants provided responses ranging in length from sentence fragments
containing less than ten words up to detailed, multi-sentence paragraphs. Organized into themes and described below, the responses suggest that UNC faculty members understand the problems associated with preserving and providing future access to research materials. In addition, the responses indicate that some researchers and research communities are proactively taking steps to address these problems.

**Figure 10 Current Needs (n=53)**

- **Server space**: 64%
- **Digital preservation services**: 11%
- **Space to store paper**: 8%
- **Accessible storage**: 17%

*Server space to store research materials in a variety of formats.* A majority of the responses (i.e., 34 out of 58) suggested that researchers will continue to rely on server space to preserve heterogeneous collections of research materials. In their responses, researchers noted the need to preserve “large databases,” “recorded measurements in ASCII format,” and “born digital information.” The responses contained many descriptive terms that suggest that researchers value their data. For example, some UNC faculty members used terms such as “stable,” “secure,” and “reliable” to describe the characteristics of the desired server space. Finally, in the words of one faculty member, the servers must provide “easy access [to the research data] when needed.”
comments also suggest the need to provide additional space to accommodate changes in the research process facilitated by advances in information technology, including a growing reliance on the use of digital images and complex scientific information technology tools.

*Storage of research materials in a location that is accessible over a specific timeframe.* Several faculty members (i.e., 9 of 58) provided specific timeframes for the preservation of their research materials. Most of the comments in this area indicate that faculty researchers store data for periods ranging from five to ten years. Some comments suggest that the timeframes may vary considerably across communities or disciplines. For example, one researcher from the humanities noted that he needs to maintain “an archive of excavation data that will still be accessible 100 years or more from now.” Another scholar from fine arts simply stated that “reports must be preserved and accessible for the indefinite future.”

*Additional space to store paper documents.* As previously noted, UNC faculty members collect, produce, and store research materials in paper formats. Responses suggest that some researchers (i.e., 4 of 58) perceive a continued need to preserve these materials, including dedicated storage locations.

*Services to support digital preservation and long-term access to research data.* Several responses (i.e., 6 of 58) indicated that researchers require advisory or technical services (including data backup) to address media, hardware, or software obsolescence. For example, one individual expressed interest in an “easy way to transfer data electronically from one form of media to the next… when the former form becomes obsolete or wears out.” Another response indicates that some researchers may have a
more pressing need for these services: “How… [do I] access data on media that are obsolete? How would I access materials on the 1970s-style mainframe datatape? 8 or 4.5 [inch] floppies? Data tape cassettes? Having access to obsolete input and output devices that will work on current systems (and are accessible to faculty users) will be a major issue in the future, and for some researchers that future is NOW.”

**Levels of Support Provided by University Organizations**

The data suggests that the majority of support received by UNC faculty members consists of storage space on local servers (70.7%). Some researchers (5) in the sciences and health noted the existence of off-site storage for paper documents and other items. However, one respondent noted that it is “nearly impossible to put things in off-site storage. If you do put it in… you have… [great] difficulty getting it back.” Just five researchers (8.6%) noted that they obtain training or have access to other support services and only two (3.4%) receive guidelines or procedures.

![Figure 11 Current Levels of Support (n=58)](image)

A third of respondents (i.e., 22 or 37.9%) used the *Other* category to provide additional information regarding the level of support received from their departments or
schools. An analysis of the unstructured responses suggests that UNC faculty members act within their communities (including partnerships with IT specialists) or on their own to bridge gaps in preservation support services. The following provides two examples of how UNC faculty members in science and health meet preservation needs:

- We purchased the server ourselves, however the department provides computer support for the server if we have a problem with it.

- We’ve purchased our own file server and run our own backup server to meet the needs of our group. We have included disk space and a backup system in a supercomputer proposal to NIH (reducing the computational ability of the machine to enable required data storage and archiving)

The additional comments also suggest how faculty members in the humanities, social sciences, and professional schools perceive the apparent lack of support in this area. For example, information provided from researchers in these fields includes short comments such as “none,” “I wish all of the above,” and “they don’t, and neither does anyone else.” In this environment, researchers act on their own to store research materials (e.g., “I simply buy external hard drives for storage at home”).

**Suggestions for Improvement**

As shown in Figure 12, the majority of faculty members identified additional space on servers as a means to improve their research data preservation needs. In describing the improved services, researchers requested “dedicated space” on “faster server[s]” to store “sensitive data.”
However, almost an equal number of faculty members responded that the university and its organization units adequately support the preservation of research materials at this time (e.g., “nothing new for now,” “I am currently served well”). Other comments include suggestions for additional space for paper documents, more advisory and technical services, improved software and hardware, and an increase in the number of workshops and training opportunities. Typical comments related to these categories include the following:

- **Biggest issue for me is the quantity of paper documents that we need to retain. This is a problem that the sponsors of research need to address.**

- **Unified plan for preservation and guidelines.**

- **Affordable backup for large desktop/workstation disks (330GB per machine) would be very helpful for our tracking and magnet control data sets, and for our confocal microscopy images.**

- **Make sure the OASIS tech person is full-time to my department and a permanent position.**
- Workshops would be very helpful.

It is interesting to note that a large number of participants offered no suggestions. Comments in this area ranged from “don’t know” and “no idea” to “I do not know enough about their current capabilities to answer [this question].”
Discussion

**Carolina Researchers Use Multiple Formats…**

Similar to other research-intensive settings, UNC faculty members collect and store large amounts of research materials in multiple formats. In addition to paper documents, this data consists primarily of research materials in digital formats. As previously noted, the digital text, images, videos and other data collected by these researchers is at risk of disappearing if they or their supporters (e.g., departments, university) do not take active steps to preserve this data. These steps include identifying the formats in use on campus and proactively monitoring technology trends to identify future developments (Smith, 2005).

The extensive use of information technology as demonstrated by the multiple digital formats used in the research process at Carolina is not limited to scientists and a select few scholars from the social sciences and humanities. Similar to other university environments, the wide-spread use of information technology crosses departmental and disciplinary boundaries. Descriptions of efforts to implement and manage institutional repositories at such places as MIT note faculty preferences for the use of data in multiple formats including text, videos, images, and datasets (Baudoin & Branschofsky, 2003). Like their counterparts at these institutions, UNC faculty members express similar preferences for a wide array of data in digital formats.
This survey also provides data that suggests that, while Carolina researchers continue to use materials in traditional formats such as paper, other non digital formats may not be as commonplace in the future. For example, only one respondent from the humanities or social sciences noted that he or she stores research data in a microfilm format. This is somewhat surprising due to the fact that recent studies of the information-seeking process of researchers from these disciplines suggest the continued high-level use of microfilm (See Brockman, Neumann, Palmer, & Tidline, 2001; Duff, Craig, & Cherry, 2004). Although the size of this sample precludes generalization, this finding perhaps indicates that the greater availability and use of digital surrogates by humanities and social science scholars has contributed to a movement away from the use of microfilm. As a result, the university must be prepared to organize, manage, and preserve an increasing amount of digital material, including digital surrogates of materials scholars traditionally preserve in microfilm format (e.g., newspapers, pamphlets, broadsides).

While use of multiple formats may facilitate the research process, it increases the complexity of the preservation process due to the large number of different formats that individuals or institutions must preserve (Wheatley, 2003). To address this challenge, a review of the field suggests that some institutional repositories recommend specific file types or identify the formats that they will or will not support (Semple, 2003; Fyffe & Warner, 2003). Efforts to change the practices of researchers when necessary to meet one of the levels of support might prove difficult. As Folk and Barkstrom (2003) note, most researchers (like other users) have significant investments in formats that they use on a frequent basis and are known within their communities, and therefore might be reluctant to change to a different format just to meet a specific policy requirement.
Institutional repositories like Michigan’s Deep Blue and MIT’s DSpace attempt to address this issue by clearly stating that, while their repositories will strive to support the multiplicity of formats submitted by researchers, they will only guarantee the preservation of a limited number of formats (DSpace Format Support Policy http://libraries.mit.edu/dspace-mit/build/policies/format.html; Deep Blue Preservation and Format Support http://deepblue.lib.umich.edu/about/deepbluepreservation.jsp). A future institutional repository at Carolina should explore the format support policies in place at Michigan, MIT, and other institutions as well as the processes they have in place to communicate with researchers about the formats desirable for future preservation.

... and Store Data in Multiple Locations

Based on survey responses, it is clear that UNC faculty members store research materials in multiple locations during the project lifecycle and after its completion. As noted in the findings, these locations include UNC supported desktops and laptops, departmental servers, portable media, AFS space, and paper. The storage of data in multiple locations during a project’s active phase suggests that UNC researchers desire ready access to their data from multiple work locations, including from their office, classroom, laboratory, home, or other off-site locations.

Information technology, especially the Internet, supports and facilitates our ability to create virtual workplaces in any location and at any time. Business consultants, health care providers, and academic researchers operate in this new environment with an increased productivity, and have come to rely on it as an inherent component of their professional life. When asked about how web-based services could support their scholarly endeavors, faculty members at the University of Rochester provided feedback
that reflects their practices and work habits in this new environment, including a desire to access their research from multiple computers, share data with colleagues, and disseminate the results of their efforts (Lindahl & Foster, 2004).

Although most research data created today is born digital (Hedstrom, 2003), a large number of UNC faculty members store the results of their effort on paper, and nearly half use portable media for this purpose. This suggests that perhaps researchers do not trust or have access to reliable long-term storage for their digital data. It could also indicate that they lack timely access to data stored on servers or in other locations. Finally, it could just indicate that researchers on campus are not aware of or informed of current options to store research materials on department or university servers.

The current practice of storing research materials in multiple locations presents a key challenge to the successful preservation of research materials. In addition to issues related to the collection of this data from multiple storage locations, this practice raises questions about how scholars currently manage access to authentic versions of their materials as well as secure them from unauthorized use. The creation of an institutional repository as a central location for the ingest, storage, and preservation of research data will help to address some of these issues.

**UNC Faculty Members Engage in Highly Collaborative Research Efforts**

Collaboration plays a key role in the research process. The practice can range from the sharing of data between individuals during a project’s formative stage to co-leadership roles for all project-related decisions, including the publication and dissemination of research results. Houghton, Steele, and Henty (2004) identified several
factors driving collaborative efforts, including pressure from project funders seeking to maximize their investment and a need to acquire specialized skills or instruments to complete project work.

As the findings suggest, UNC-Chapel Hill faculty members frequently participate in collaborative efforts. Their collaborators may work in the office next door, across campus, or even overseas. A high-level of collaboration may indicate a need to store data at UNC and provide broad access to this data. To meet this need, the university may need to create new information technology tools to facilitate this process and ensure that the data remains in a safe, secure location.

*An Opportunity Exists to Provide Additional Support for the Preservation of Research Data*

As the evidence suggests, UNC faculty members are highly innovative researchers who understand the implications of working with digital data, and are particularly concerned with preserving it for future use. Their comments about their preservation needs suggest that, while they receive some support from their department or school, this support is limited to storage space on a server and often on back-ups. As noted in the literature, this single service is not enough to guarantee the survival of digital data for any length of time, let alone the 100 year timeframe requested by one scholar. Commitments of this length require sustained support from a stable institution. Lynch (2003) points out the futility of individual faculty members in trying to establish and maintain stable preservation environments on their own. In most cases, they simply lack the time and technical expertise to monitor digital data, migrate it to new formats, manage version control, and secure it from unauthorized access and use. Collaborative
efforts that cross departmental, regional, and even national boundaries increase the complexity of these tasks.

In regard to questions related to long-term preservation needs, the fact that so many individuals answered “don’t know” or used words with similar meanings suggests that the university and its information stewards must improve or expand on training and communication related to digital preservation. In addition to final publications and project reports, digital research materials represent intellectual assets worthy of preservation for future generations. Like other digital content on campus (e.g., learning objects), a growing number of universities recognize the need and their responsibility to preserve these valuable materials (Hey, 2004; Pritchard, Carver, & Anand, 2004; Lynch, 2003; Rogers, 2003). Institutional repositories provide universities with an ideal infrastructure to disseminate this content throughout the scholarly community (Branin, 2004).

As previously noted, issues associated with digital preservation are not unique to the UNC research environment—these issues impact organizations of all types and sizes, including governments, multinational corporations, and universities. In most university settings, faculty members lack the knowledge to understand issues related to digital preservation at a level of detail sufficient to make preservation-friendly format decisions (Spindler, 2004). While organizations can develop policies, procedures, and workflows to facilitate and support digital preservation activities, they must take additional steps to ensure that individuals understand the complexities of the crisis and can actively participate in addressing it from an informed perspective. Successful digital preservation efforts include informed data producers who recognize that their obligations to digital
preservation start at the point of creation. This represents a departure from the traditional records lifecycle in which data producers and in some cases records managers took responsibility for safeguarding data during its active life. Archivists and other data stewards responsible for long-term preservation typically did not take responsibility for the records until the producer submitted the data to the archive or data center at the end of its active life. Due to the fleeting nature of digital materials, information stewards simply cannot wait until the end of the traditional lifecycle before becoming active participants in the preservation of these materials (Hedstrom, 2003; Hedstrom & Ross, 2003).

In a university setting, communication, education, training and other interactions between information producers and those who manage and preserve their data play a key role in enabling institutions to proactively engage faculty members in preservation efforts. Education, for example, plays a critical role in establishing strong, collaborative processes between archivists and data producers. Archivists, for example, must clearly convey the benefits of archiving data to researchers. This responsibility includes the dissemination of best practices and standards that promote and support the reuse of research data (Crabtree & Donkowski, 2006).

Strong partnerships involving data producers, information stewards, and other community members is just one example of a best practice that the university should seek to emulate in its current efforts. Successful repositories rely on partnerships with stakeholders to ensure that its services meet their needs (Lynch, 2003; Chavez, Crane, & Sauer, 2006). Partnerships provide a useful vehicle for communication. They enable repository administrators and managers with an opportunity to obtain detailed knowledge of the practices and habits of their stakeholders, especially data producers, and the
problems they face managing information. Information stewards and data curators can also use communication channels opened through partnerships to educate data creators of the importance of selecting supported formats and creating descriptive metadata.

**Implications for Current Efforts to Develop an Institutional Repository**

The results of this study provide the university and the Digital Curation / Institutional Repository task force with useful baseline data about the research collection and storage practices of UNC faculty members. As the literature suggests (e.g., Smith, 2005; Barton & Waters, 2005), collaborating with key stakeholders of and contributors to an institutional repository is a crucial first step. In laying out a possible roadmap for preservation of digital research materials, Smith (2005, 18) notes that institutions must “discover who on campus has research data, in what forms, with what use requirements, and with what current options for archiving.” The results noted in this paper provide a first step in obtaining this information and identify some potential challenges that the university and task force potentially face during this process. The university can use the data to open a dialog with its research data producers and as a key input into the development of a larger university-wide study designed to describe the research process on campus in more depth.

In the final analysis, the types of responses provided by faculty members throughout the survey clearly suggest that other factors and concerns related to their larger role in the academic community impact their ability to collect, manage, and preserve information. As one researcher noted below, the ubiquitous nature of information in the Digital Age and the technologies that facilitate its creation and
dissemination can overwhelm even individuals who were at one time technologically savvy.

*Once upon a time, I might have been up to date with the bells and whistles of different software, but those times are long, long gone, and I now need a tech person to help me with accessing storage technology among other things. I just don’t have time, along with lacking the expertise, to manage my backlog of emails, let alone articles, data, images, etc. So they sit on my hard drive or in the server, sometimes even when I don’t really need them [ever] more.*

In this environment, perhaps the greatest service that the university can offer its faculty members is a set of policies, procedures, and technological infrastructure that supports and sustains their research efforts. An institutional repository, designed to meet their specific research needs may serve to fill this role and provide a stable environment for the long-term preservation of research data.
Conclusion

This study explored how UNC faculty members collect, store, and preserve research data. Its results suggest that the research process at Carolina includes the use of rich sets of research materials in a variety of formats, including text, video, images, and datasets. Researchers store this material in a variety of formats in multiple locations, and share this data with collaborators located on campus, across the nation, and even overseas. Although faculty members acknowledged receiving preservation support from their departments and the university, they also identified several areas for improvement, including additional storage space on servers and the need for new technical services.

While the study’s findings present a useful baseline of current practices, the need exists for subsequent efforts to obtain more rich information via semi-structured interviews and demonstrations. Questions raised in this study that future efforts should address include but are not limited to the following:

- In the case of storing documents in multiple locations, how do researchers know which document is the correct or most authoritative version? How do researchers protect or secure data from access by unauthorized individuals?
- What steps do UNC faculty members take to ensure that other researchers can access and reuse their data in the future? What types of metadata (if any) do they assign to the documents?
What steps or processes are in place to migrate data to new formats? Do researchers actively monitor data stored on hard drives, portable media, or on department/university servers?

What steps do researchers take to manage data exchanged via email? Do they leave data in their inbox (including attachments) or move the information to the appropriate folder?

Is the current environment sustainable over the long-term? Do costs associated with developing and implementing an institutional repository for digital materials outweigh the risks of maintaining the status quo?

The data gained from answering these questions and others will provide future implementers of an institutional repository at Carolina with additional insights into faculty research data practices, needs, and expectations. The university should share the results of this research with other U.S. and international universities to disseminate potential best practices and to open doors to possible future partnerships or collaborative efforts.
References


Retrieved November 2, 2006, available from


http://uksg.metapress.com/media/m3t1fan03l2xrjfb3r8x/contributions/r/n/w/q/rnw qyqj54pvvxyun.pdf


http://www.lib.ed.ac.uk/sites/digpres/dpreport.pdf


Appendix A

Survey Instrument

Please answer the following questions based on your research experiences over the last three years.

1. What types of data do you collect during the research process? (Please check all that apply.)

- Observations or measurements
- Documents or texts
- Images or films
- Artistic or cultural artifacts
- Other

If other, please describe:

2. In what formats do you store research data? (Please check all that apply.)

| Text (e.g., HTML, Text, Rich Text Format, XML, SGML, ASCII) | Video (e.g., MPEG, QuickTime) |
| Image (e.g., GIF, JPEG, PNG, TIFF, BMP, Photo CD) | Applications (e.g., Science, GIS, Mathematics and Statistics, Visualization) |
| Audio (e.g., AIFF, MPEG, RealAudio, WAV) | Databases (e.g., MS Access, MySQL) |
| Adobe PDF | Microfilm |
| Other |

If other, please describe:

3. Where do you store research data during its active life? (Please check all that apply.)

| On the hard drive of a UNC-supported desktop | On AFS space |
| On the hard drive of a non UNC-supported desktop | On other campus-wide computer storage |
| On the hard drive of a UNC-supported laptop | On a department/school server |
| On the hard drive of a non UNC-supported laptop | On portable media (e.g., CD/DVD, tape, floppies, zip drives, jump drives) |
4. Where do you store research data after the completion of your research endeavor or project? (Please check all that apply.)

- On the hard drive of a UNC-supported desktop
- On the hard drive of a non UNC-supported desktop
- On the hard drive of a UNC-supported laptop
- On the hard drive of a non UNC-supported laptop
- On paper hardcopies
- On AFS space
- On other campus-wide computer storage
- On a department/school server
- On portable media (e.g., CD/DVD, tape, floppies, zip drives, jump drives)
- Other

If other, please describe:

5. For what purposes do you keep research data past its active life? (Please check all that apply.)

- Potential reuse of data (personal)
- Potential reuse of data by other researchers
- Public relations / outreach
- Required by publishers
- Required by funding agencies (government, industry, or private foundations)
- Required by the university
- Other

If other, please describe:

[Note to IRB reviewers: subjects will answer one of the three questions that follow (i.e., A, B, or C) depending on how they answer question 5.]

A. What types of data do funding agencies require you to store after the completion of a research endeavor or project? (Please check all that apply.)

- Administrative records
- Research materials (e.g., images, datasets, lab books, field notes)
- Reports / presentations
- Articles / publications
- Correspondence
- Web sites / portals
If other, please describe:

B. What types of data do publishers require that you store after the completion of a research endeavor or project? (Please check all that apply.)

<table>
<thead>
<tr>
<th>Administrative records</th>
<th>Research materials (e.g., images, datasets, lab books, field notes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports / presentations</td>
<td>Articles / publications</td>
</tr>
<tr>
<td>Correspondence</td>
<td>Web sites / portals</td>
</tr>
<tr>
<td>Public relations / outreach</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:

C. What types of data does the university require you to store after the completion of a research endeavor or project? (Please check all that apply.)

<table>
<thead>
<tr>
<th>Administrative records</th>
<th>Research materials (e.g., images, datasets, lab books, field notes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports / presentations</td>
<td>Articles / publications</td>
</tr>
<tr>
<td>Correspondence</td>
<td>Web sites / portals</td>
</tr>
<tr>
<td>Public relations / outreach</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:

6. Do you collaborate with other researchers?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

[Note to IRB reviewers: subjects will answer the four questions that follow (i.e., A, B, C, D) if they answer yes to question 6.]

A. Where are the other researchers primarily located?

<table>
<thead>
<tr>
<th>In my department / school</th>
<th>In other UNC departments / schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>At other U.S. academic institutions</td>
<td>At other international academic institutions</td>
</tr>
<tr>
<td>At U.S. public sector institutions</td>
<td>At U.S. private sector institutions</td>
</tr>
<tr>
<td>At U.S. non profit institutions</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:
B. How do you share research data? (Please check all that apply.)

<table>
<thead>
<tr>
<th>Method</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person to person</td>
<td>Paper</td>
</tr>
<tr>
<td>Email</td>
<td>UNC department / school server</td>
</tr>
<tr>
<td>Portable media (e.g., CD/DVD, tape, floppies, zip drives, jump drives)</td>
<td>Other non UNC server</td>
</tr>
<tr>
<td>Web interface / portal</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:

C. Where do you store research data at the conclusion of collaborative efforts?

<table>
<thead>
<tr>
<th>Location</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just at UNC</td>
<td>Just at other site</td>
</tr>
<tr>
<td>At UNC and other sites</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:

D. If at UNC, where do you store research data after the completion of your collaborative research endeavor or project? (Please check all that apply.)

<table>
<thead>
<tr>
<th>Storage Location</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>On the hard drive of a UNC-supported desktop</td>
<td>On AFS space</td>
</tr>
<tr>
<td>On the hard drive of a non UNC-supported desktop</td>
<td>On other campus-wide computer storage</td>
</tr>
<tr>
<td>On the hard drive of a UNC-supported laptop</td>
<td>On a UNC department/school server</td>
</tr>
<tr>
<td>On the hard drive of a non UNC-supported laptop</td>
<td>On portable media (e.g., CD/DVD, tape, floppies, zip drives, jump drives)</td>
</tr>
<tr>
<td>On paper hardcopies</td>
<td>Other</td>
</tr>
</tbody>
</table>

If other, please describe:

7. What are your long-term data preservation needs?

[Open ended Question]

8. How do your department, school, or other university units help meet these storage or preservation needs? (Please check all that apply.)

<table>
<thead>
<tr>
<th>Support Services</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides storage space on local server</td>
<td>Distributes guidelines and procedures</td>
</tr>
<tr>
<td>Provides off-site storage for paper</td>
<td>Offers training and other support services</td>
</tr>
<tr>
<td>documents and other items</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--</td>
</tr>
<tr>
<td>□ Other</td>
<td></td>
</tr>
</tbody>
</table>

If other, please describe:

9. How could your department, school, or university better meet your research data preservation needs?

[Open ended Question]

10. Please identify your school / department. (In the case of multiple appointments, please select your primary school / department.)

<table>
<thead>
<tr>
<th>Academic Affairs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>College of Arts and Sciences</td>
<td></td>
</tr>
<tr>
<td>□ Fine Arts and Humanities</td>
<td></td>
</tr>
<tr>
<td>□ Sciences</td>
<td></td>
</tr>
<tr>
<td>□ Social Sciences</td>
<td></td>
</tr>
<tr>
<td>Kenan-Flagler Business School</td>
<td></td>
</tr>
<tr>
<td>School of Education</td>
<td></td>
</tr>
<tr>
<td>School of Government</td>
<td></td>
</tr>
<tr>
<td>Information and Library Science</td>
<td></td>
</tr>
<tr>
<td>Journalism and Mass Communication</td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td></td>
</tr>
<tr>
<td>Social Work</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health Affairs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentistry</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td></td>
</tr>
<tr>
<td>Nursing</td>
<td></td>
</tr>
<tr>
<td>Pharmacy</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Units</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute of Marine Sciences</td>
<td></td>
</tr>
<tr>
<td>Carolina Environmental Programs</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

If other, please describe:
Appendix B

*Invitation to Participate in the Study (Sent via Email)*

You are one of 200 faculty members selected randomly to participate in a study of the digital storage practices and needs of UNC faculty related to the research process. The knowledge gained from this study will allow the university to better support faculty research efforts.

Dr. Helen Tibbo, Professor, and Mr. John C. Schaefer, Graduate Student, both of the School of Information and Library Science at UNC-Chapel Hill are conducting this study.

**What is the purpose of this study?**
This study will survey Academic and Allied Health faculty at the University of North Carolina at Chapel Hill (UNC-CH) in an effort to document current practices related to the storage of research data. The online survey includes questions regarding digital storage and preservation needs as well as specific challenges faced by faculty in these areas.

**What will happen if you agree to take part in this study?**
You will be asked to complete a short, online survey. We anticipate that it will take you less than ten minutes to complete the survey.

**How will your privacy be protected?**
All survey input is anonymous. The only personal information being requested is the school or college in which the participant works collected as part of this study.

**What if you have questions about your rights as a research participant?**
You are free to refuse to participate or to withdraw from this study at any time without penalty and without jeopardy.

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.

**How do I access the survey?**
If you agree to participate in this study, you can access the survey by clicking on this link [Insert link here]. You can stop your participation in the survey at any point without penalty or jeopardy.

This link will remain active until [Insert date here].

If you have any questions or concerns about participating in this study, please contact
John C. Schaefer at jschaefc@email.unc.edu.

Thank you for considering our request.