

Samantha A. Guss. Assessing the Effects of Institutional Review Boards on Social Science Data Archiving in Digital Repositories. A Master's Paper for the M.S. in L.S. degree. April, 2009. 61 pages. Advisor: Helen Tibbo.

This study addresses the intersection of digital repositories and archives and institutional review boards (IRBs) with regard to opinions on data reuse from the IRB members themselves and researchers in social and behavioral science disciplines whose research requires IRB approval. The issue was examined from several perspectives, utilizing a survey of social and behavioral science researchers, interviews with IRB members, and an analysis of the language used in blank Informed Consent forms from previously approved studies related to potential data reuse. Combined, the data gathered provides multi-faceted insight into these two groups, especially regarding IRB application language and subjects' informed consent. It recommends early intervention by repositories, open discussion within Institutional Review Boards on data archiving in social and behavioral science, increased guidance by IRBs and repositories for applicants, continued education, and increased transparency and communication among IRBs and researchers.

#### Headings:

Digital preservation

Electronic data archives

Institutional repositories

Professional ethics

Social science research

ASSESSING THE EFFECTS OF INSTITUTIONAL REVIEW BOARDS  
ON SOCIAL SCIENCE DATA ARCHIVING IN DIGITAL REPOSITORIES

by  
Samantha A. Guss

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

April 2009

**Approved by**

---

Helen Tibbo

## Table of Contents

1. Introduction.....	3
2. Literature Review.....	6
2.1 Data Reuse in Social Science.....	7
2.2 Tension Between Researchers and IRBs.....	10
2.3 Institutional Repositories and Data Archives.....	12
2.4 Outreach and the Librarian's Role.....	14
2.5 Summary of Literature.....	16
3. Method.....	18
3.1 Part A: Survey.....	19
3.2 Part B: Interviews.....	20
3.3 Part C: Content Analysis.....	21
4. Results.....	23
4.1 Part A: Survey.....	23
4.2 Part B: Interviews.....	35
4.3 Part C: Content Analysis.....	39
5. Discussion.....	43
5.1 Limitations.....	48
5.2 Recommendations.....	49
6. Notes.....	52
References.....	53
Appendix A: Survey Questions.....	55
Appendix B: Interview Instrument.....	58

## List of Figures

Figure 1: Survey respondents' academic status.....	23
Figure 2: Survey respondents' school or department primary affiliation.....	24
Figure 3: Distribution of coded "data management plans".....	26
Figure 4: Answers to Question 5: "Did you follow, or are you following, your data management plan as outlined in your IRB application?".....	27
Figure 5: Answers to Question 6: "Based on your last application's wording, as it was approved by the IRB, would you feel comfortable archiving the de-identified data resulting from your study in a digital repository?".....	28
Figure 6: Answers to Question 7: "Have you ever archived data from a study in a digital repository?".....	30
Figure 7: Answers to Question 8: "Would you consider doing so [archiving your study data in a digital repository]?".....	31
Figure 8: Answers to Question 12: "Do you think the IRB would approve a study that included a plan to archive de-identified data for future use by you or other researchers?".....	33
Figure 9: Distribution of coded blank Informed Consent forms.....	40

## **1. Introduction**

This study addresses the intersection of digital repositories and archives and institutional review boards (IRBs) with regard to opinions on data reuse from the IRB members themselves and researchers in social and behavioral science disciplines whose research requires IRB approval.

Digital repositories, including institutional repositories as well as disciplinary archives, are difficult to define and can encompass many things, but at the most basic they are institution-based digital archives in which digital scholarly output of many types is curated over time and made available publicly. From a resource for teaching materials to a new kind of scholarly publishing, IRs have a variety of potential uses. One area of particular interest is for institutional repositories or other archives to acquire the raw data sets generated from research studies or by researchers. This data would be maintained in usable form so that later researchers could either verify the data reported in a resulting article or reuse the data in a different way for their own study. Few researchers have yet to embrace the use of institutional repositories. One potential reason for this is confusion among researchers as to what is allowed concerning open access to data and reports. No IR literature has addressed IRBs yet.

Institutional review boards preside over research involving human subjects at a given institution (such as a university or hospital), and are federally mandated to ensure that those human subjects are protected from any unnecessary harm resulting from the study. The Office for Human Research Protections, a division of the United States

Department of Health and Human Services, oversees institutional review boards and provides resources, but individual IRBs operate independently and create their own procedures and documentation. All IRBs exist to uphold professional research ethics, as outlined in the Belmont Report (1978): Respect for Persons, Beneficence, and Justice. Practical applications charged to ethical researchers and IRBs are Informed Consent, Assessment of Risks and Benefits, and Selection of Subjects. Large institutions often have several review boards, responsible for applications from different departments or disciplines. IRBs have existed in the United States since 1974.

Researchers must apply for IRB approval before a study has started, outlining specifically what will be done and to whom, and the IRB decides whether the study is ethical and may be conducted as written or more commonly, what needs to be changed to make the study more clearly ethical. Data resulting from social and behavioral research studies sometimes contain confidential information about the subjects studied; IRBs are very interested and often quite strict about how the researcher protects this data from becoming public. In order to be shared, this data usually must be de-identified so that it cannot be linked back to the individual who provided it. IRB applications differ slightly among institutions, but the intent is the same. Since applications are lengthy and approval often requires revisions (taking time they might prefer not to spend), researchers must be careful and deliberate in documenting their intended actions, including data management.

When framed this way, it might seem that digital repositories and archives collecting social science data are somehow in opposition to the goals of the IRB process. However, the data collected in this study provides evidence that they are not truly at odds and can co-exist to the benefit of everyone. This study examines IRB members'

perceptions of social and behavioral science data archiving in addition to researchers' perceptions of what the IRB will or will not allow them to do. Increasing the transparency of the process helps to remove obstacles for researchers' contributions to digital repositories, furthering their development and use.

From reviews of the various literature informing this issue, the following research questions were formulated to guide this study:

- What are IRB members' perceptions and knowledge of digital repositories and the practice of data archiving in social and behavioral science?
- What guidelines and procedures do IRB members think should be followed when making social and behavioral science research data available in a repository?
- How do IRB considerations affect a researcher's decision whether to archive his/her data in a digital repository?
- If IRB considerations are indeed a barrier to archiving of social and behavioral science data, how can the barrier be removed?

## 2. Literature Review

It is suspected that a variety of factors, including researchers' perceptions of IRB procedures, is a factor impeding faculty participation in digital repositories and other data sharing venues. A review of the literature reveals that this topic has not been directly addressed before, providing justification for an exploratory study such as this one.

IRBs' intentions and procedures, and even their very reason for existence, seem to be widely misunderstood or ignored. IRB applications are often seen as no more than an annoying extra step requiring time and effort better spent elsewhere and the IRB itself as inefficient, overly strict, or wanting to cause headaches for researchers. Speed is widely cited as a source of contention with IRB applications and, at least anecdotally, it seems as though researchers are interested in getting their applications approved as fast as possible—and will fill them out with this in mind. For social science data archiving, this practice could be harmful: if researchers suspect faster approval for applications that propose destroying data, digital repositories will never have a chance to recruit this data for their collections.

Disciplinary data repositories for social science have existed for close to a century while institutional repositories are a much newer endeavor, but they share many of the same challenges related to collection development. Institutional repositories have been slow to catch on for a variety of reasons and a sizeable body of literature exists contemplating these reasons and ways to address them. Overall, removing obstacles and educating librarians and researchers alike will contribute to the growth of these



repositories, which is arguably beneficial to all of the stakeholders involved. The librarians and archivists responsible for the creation and upkeep of the repository will have to take an active role as liaisons between the repository and its contributors: the data creators. For new repositories, this will include initial recruitment of materials as well as ongoing outreach activities, both of which require open communication. To do this, they will need as much information as possible to answer researchers' questions and instill confidence. This study is meant to add to this body of useful information.

The literature related to this question spans social science disciplines like sociology, psychology, and communication studies, ethics, studies of IRBs themselves, as well as library and information science literature on digital repositories and librarians' roles in facilitating faculty research.

## **2.1 Data Reuse in Social Science**

The idea of reusing study data in social science disciplines is not a new idea and data archives, especially for quantitative studies, have existed for many decades. The H.W. Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill was founded in 1924, the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan was established in 1962, and the Roper Center Public Opinion Archives at the University of Connecticut was founded in 1947 with data going back to the 1930s. Internationally, the Council of European Social Science Data Archives (CESSDA) was created in the 1970s, bringing together individual archives, many of which have existed for much longer.

Although analysis of secondary data has not become commonplace, researchers in social science have been arguing for its value for at least thirty years. Finifter (1975) was one of the earliest to outline the benefits of storing data for later use. This article cited various uses ranging from simple re-purposing and validity/methods testing to multiple methods triangulation by later researchers. Even in the mid-seventies, well before the birth of the Web, Finifter envisioned a system of alerting and sharing with other researchers and argued that preserving the data is useless if no one knows about it or can access it.

There are problems, however, with long term preservation of social science research data particularly related to ensuring participants' confidentiality and privacy. Professional research ethics call for informed consent--the subject must explicitly agree to participate, knowing what will happen in the study and their risks for participating--and careful protection of any personal details collected during the study that would allow a third party to trace responses back to individuals. Both of these things can be problematic for data archiving. When many social science data archives began, there were no institutional review boards at all. Even in the last ten years, IRB applications have become longer and more detailed. The IRB application at the major research university studied here is approximately fourteen pages long. Compare this with a filled-out, approved application from 2004 from the same university that totals four pages.

Parry & Mauthner (2004) argued that subjects sign informed consent agreements only for the study at hand: they did not consent to other studies associated with secondary use of that data, pulled from an archive by a different researcher perhaps years later. In addition, the authors discussed the difficulty of de-identifying qualitative (versus

quantitative) research data. Instead of deleting a few columns of personal data from a spreadsheet, it might involve censoring interview documents line by line, still possibly missing identifying pieces. In addition, according to Parry & Mauthner, "ensuring anonymity might require removing so much detail that the data are rendered meaningless." This is not a complete impasse, but social scientists such as Parry & Mauthner have not been satisfied so far with the recommendations and assistance with de-identification for qualitative studies.

Scott (2005) also addressed issues of subject identification as it relates to communications research, especially focusing on anonymity (never recording personal data about participants at all) and confidentiality (carefully guarding any personal data that is recorded). The study did not argue for anonymity but warned that researchers and IRBs should carefully consider whether collection of personal information is necessary considering the difficulties of keeping it absolutely private. In addition, the author discussed the importance of researchers cultivating relationships with members of the IRB to encourage open communication on these and other issues.

Easter, Davis, & Henderson (2004) discussed the issue of confidentiality in survey research as new technology makes it difficult to ensure that participants cannot be traced to their data, especially when more than one party is involved in the research (as is often the case in social science endeavors). They, too, suggested that researchers and IRB members think carefully about what personal data absolutely needs to be collected and advocated open communication and education for both on data protection issues. Additionally, they recommended a system that allows for reassessment and easy modification of a researcher's data management plan.

## 2.2 Tension Between Researchers and IRBs

Another important issue impacting social science research that must be considered is the tension between researchers and IRBs, which is probably better described as researchers' animosity or annoyance for the IRB process. The American Psychological Association (APA) created a task force in 2007 specifically to address this issue that persists although, as the report argued, the IRB and researchers do ultimately have the same goals (APA Presidential Task Force on Institutional Review Boards and Psychological Science, 2007). The task force reported that one reason for this tension is the perception that biomedical IRB procedures are unnecessarily applied to social science researchers. One of its recommendations to alleviate the tension was to expand education on ethics, the purpose of IRBs, and submission of IRB applications at all levels, beginning in undergraduate psychology programs.

In an empirical study of social and behavioral researchers, Ashcraft & Krause (2007) used a survey and open ended questions to learn about their subjects' experiences with IRBs. Their survey results revealed a "surprising" amount of satisfaction with IRB application processes, with nearly half claiming that their IRBs had a "reasonable" turnaround time. However, the comments they received were largely negative or accounts of IRB "horror stories." This study also found that, as expected, timeliness is a major source of tension for researchers regarding IRBs and lack of time was cited as a major reason among those who chose to bypass IRB approval. Much was revealed about researchers' perceptions of their IRBs as well: for instance, many felt that their applications were approved quickly because their methods presented no real risk to participants, but others gave credit to efficient IRBs for the timeliness, and some believed

that their personal reputation with IRB members was to their advantage. Many also claimed that over the years they had learned to write their applications with an eye to expediency. The authors found that several researchers questioned the need for IRBs at all, and responded in their article with justification and a call for increased advocacy and education about research ethics boards.

Along the same lines, Candilis, Lidz, & Arnold (2006) emphasized that the lack of research on how IRBs function and make decisions leads to confusion, lack of transparency, and questions about IRBs' necessity. They especially highlighted that IRBs should be studied so that they can be more easily reformed and so that more stakeholders can have a say in how that reform takes place. With more information and more input, they argued, the researchers will not feel as frustrated by the process.

In an article specifically addressing IRB-researcher tension, Fitch (2005) provided recommendations to improve the situation for both sides. Among these were promoting open communication between the researchers and the board and asking (and preparing to answer) questions about applications to make sure everyone is absolutely clear. It was suggested that IRB members follow directions consistently and that researchers make an effort to realize the complexity of an IRB's process as well as its importance, even when it is not immediately apparent on a case-by-case basis. Perhaps most importantly, Fitch recommended that social science researchers stop thinking that institutional review boards are malicious and out to make their work more difficult.

Almost all of the literature in this category recommends education and openness between IRBs and researchers seeking their approval as a mechanism to alleviate tension. This might be a speciously easy answer to a complex problem, but it is worth noting.

### 2.3 Institutional Repositories and Data Archives

While social science data archives have existed for many years, institutional repositories are a relatively new venture for libraries and archives trying to capture and preserve their university's output and assist researchers themselves. Institutional repositories are difficult to define<sup>1</sup> and can encompass many things, but at the most basic are institution-based digital archives in which digital scholarly output of many types is curated over time and made available publicly. For social science research, this output could consist of things like data sets or interview transcripts generated by research studies. Institutional repositories, in addition to other types of data archives, would preserve this data in usable form so that later researchers could use the data as Finifter (1975) and others suggested. Despite their potential benefits, IRs haven't been quick to catch on among scholars; studies of reasons for use and non-use are useful in discussing the IRB's role in social science data archiving.

Kim (2007) studied faculty to see what output they made public and why, their contribution to IRs, and factors that make them contribute or not. The sample obtained in the study did not include any contributors, so Kim learned more about non-contributors and their motivations. Overall, knowledge about IRs was low, but faculty who planned to contribute in the future were convinced of a variety of benefits, including larger dissemination of their work and "altruistic impetus for making research work available to other researchers." In addition, Kim found many to be confused about copyright issues and whether they would be allowed by their publishers to contribute.

Davis & Connolly (2007) also studied faculty non-use of an institutional repository and their findings pointed to a variety of reasons that are important for digital

repository promoters to take into consideration. Like Kim, they found that a lack of knowledge about institutional repositories and their benefits was a big reason it was not being utilized. In addition, various disciplinary cultures were not supportive of sharing unfinished work and without any rewards system or recognition for deposit, the faculty surveyed were not eager to do anything requiring extra work. Davis & Connolly's findings also suggest that the inertia associated with the traditional scholarly output model caused non-use; the authors discussed the risks associated with being an early adopter of new technologies in relation to the way faculty are rewarded for their research.

The MIRACLE (Making Institutional Repositories in A Collaborative Learning Environment) Project, funded by the IMLS, has employed surveys, phone interviews, and case studies to investigate the status of institutional repositories in the United States, their infrastructure and collections, and their users and potential users. Aside from numerous findings regarding IR use and implementation, the MIRACLE study also confirmed the findings of others that recruiting digital content for IRs is difficult (2007). This study also found that "staff involved with the IR effort have voracious appetites for information about IRs," highlighting the importance of continued research and documentation on IRs.

Foster & Gibbons (2005) conducted an anthropological study of faculty, hoping to illuminate the non-use problem by determining what faculty *do* want. Their findings inform both IR development and how librarians and archivists should market the IR to encourage use. Most relevant to social science research and IRB members approving this research are Foster & Gibbons' findings that faculty do want to make their work available to others and keep up in their fields and are interested in easy version control, collaboration, and want someone else to take responsibility for upkeep of servers and

digital tools (all of which such digital repositories can fulfill). They are also concerned with ownership and security of their materials and want to be careful not to violate copyright. In other words, the study's findings suggested that faculty want to play by the rules, but want things to be easy that are not directly related to their own scholarship. The authors also suggested that marketing and outreach of IRs will have to take faculty desires into consideration in order to be successful.

## **2.4 Outreach and the Librarian's Role**

Creating open and productive relationships between IRBs, digital repositories, and social science researchers seems as though it will require someone to act as a liaison, understanding the processes of all three and helping the entire research cycle operate smoothly--a natural role for librarians. In a content analysis study of deployed or "live" institutional repositories, McDowell (2007) found evidence that IR use is not growing significantly; more specifically, that faculty output does not just "find its way" into IRs as some early adopters had hoped. The discussion of these findings emphasizes that IRs are not alleviating any kind of scholarly publishing crisis at this point. This too suggests that future use of IRs and other digital repositories will require active prompting.

Watson (2007) also examined faculty perceptions of IRs, including reasons for non-use and hesitation about use and found that faculty were largely unaware of IRs but saw some benefit to them when they were explained. In addition, the author found that researchers preferred to let others take care of the technical aspects of deposit and that they viewed depositing materials as extra work in which they were not very interested.



The study concluded that the appropriate next step is to integrate depositing of data into the research process, therefore being involved early on in the data cycle.

Green & Gutmann (2007) also discussed early involvement in research as a critical step in recruiting material to institutional repositories. According to the authors, librarians and archivists should forge partnerships throughout the data lifecycle, providing assistance and outreach to researchers as the data is created. This intervention would work towards the goal of data created with best archiving practices in mind, in addition to making researchers aware of archiving their materials in digital repositories.

The MIRACLE study (Markey et al., 2007) found that librarians are more often IR developers and stewards than anyone else on campus and that IRs help the libraries build new relationships. Responses to their survey suggested that the collaboration with other entities like campus archives, student services, and digital asset management systems, just to name a few, necessary to implement the IR lead to "increasing the library's role as a viable partner in the research enterprise."

Gierveld (2006) discussed another method of outreach: applying the concept of Social Marketing, which focuses on changing a group's behavior for the good of society, to promotion and outreach for institutional repositories. In this piece, the IR is envisioned as a product and its attributes and benefits are laid out in detail along with strategies to promote each--essentially taking a marketing approach to "selling" institutional repositories. Gierveld rationalized that this approach is appropriate because it is driven by the technology available and visionaries who are not directly related to its use; it was not a response to an articulated need by its target group. Therefore, a successful marketing strategy might be similar to one promoting public health behavior. Gierveld's article

focused on making the users central to the IR's development and eventual importance. Naturally it also requires someone to administer the marketing efforts.

In a review of the literature on institutional repositories, Allard, Mack, & Feltner-Reichert (2005) determined that analysis of librarians' roles in IRs has not been sufficiently studied and more research is needed in this area. They found that institutional repository topics like defining IRs, implementation, management, and outcomes are well represented in the professional discourse but outreach, a natural role for the librarian, needs more attention in order for IRs to succeed.

It is important to note, however, that librarians should be liaisons encouraging a wide range of digital repository uses and helping researchers to negotiate the process, but not be gatekeepers, as Lynch (2003) argued in an early piece on institutional repositories. He envisioned IRs as infrastructure for a new kind of scholarly publishing that responds to the needs of its users rather than mandating what it will or will not support. This too is a natural role for librarians and archivists, who already increasingly employ user-centered practice.

## **2.5 Summary of Literature**

From a review of the literature on this topic, several things are apparent. For one, there seems to be a great deal of tension between institutional review boards and the social science research community whose work requires their approval. This tension appears to be caused by lack of transparency, communication, and lack of information about IRBs' roles or how these could be modified or improved. The literature about promptness of IRBs and the various problems with de-identification of social science data

suggests that researchers in these fields might think that promising to destroy their data is either necessary for quick approval or preferable to IRB members reviewing their applications. It is also likely that these researchers give little or no thought to archiving their study data at all at this stage, which possibly affects their participation in institutional repositories or data archives after the fact. Meanwhile, very little is actually known about IRBs' opinions or preferences regarding social science data archiving.

Generally, there is not sufficient literature at this time to explain the relationship between IRBs, social science researchers, and digital repositories, making this an important topic to study. This study is mainly exploratory, but seeks on a broader level to add to this body of knowledge, addressing the issues outlined here and possibly inspiring a variety of practical outcomes.

### **3. Method**

The evidence gathered by this study is meant to clarify a piece of the IRB process in order to assist researchers applying for IRB approval, support those recruiting content for digital repositories, and inspire discussion within IRBs about issues of data archiving. In order to accomplish these goals, both groups involved in the IRB application and approval process were studied: the social and behavioral science researchers submitting applications to an IRB and the institutional review board members themselves. From the researchers, this study sought to learn about past IRB application behavior, opinions on data archiving, and perceptions about the IRB and its process. From IRB members, it probed general impressions on data archiving (especially related to ethics), past encounters with researchers' data management plans, and how researchers might best incorporate and execute data archiving plans. In addition to these survey and interview responses, a simple content analysis was employed to further investigate researchers' and IRB members' behaviors.

Because the research questions presented here are exploratory and rely on the opinions (and actions) of two separate populations, a combination of methods was deemed most appropriate to collect useful data. A large sample of social and behavioral science researchers completed an online survey, while several members of a Behavioral IRB at a large research university were interviewed via telephone. The content analysis was carried out using Informed Consent forms from past Behavioral IRB-approved studies. These forms, hereafter called "blank Informed Consent forms," were obtained

from past approved IRB applications: they are the exact forms researchers used to attain consent from their subjects, but have not been signed or dated by anyone.

### **3.1 Part A: Survey**

The first part of the study was a survey of social and behavioral science researchers at a large research university to collect data about their interactions with the IRB related to data management plans and data archiving. A survey is a particularly apt instrument to probe this population because it is relatively large, scattered, and diverse, meaning that it would be difficult to gather a representative sample using many other methods.

Using a report generated by IRB administrative staff at a large research university, the sample of social and behavioral researchers was identified according to which departments submitted the most Behavioral IRB applications in 2008. Email addresses were gathered from these departments' websites (with help of the university-wide directory when necessary), including all listed faculty, PhD candidates, and Master's students from each department, and excluding adjunct faculty. A link to the survey, created with Qualtrics survey software, was emailed to each of these addresses. Those who had not responded within a week of the first email were sent a reminder, also containing a link to the survey.

Subjects were asked a combination of open and closed questions (see Survey questions in Appendix A), designed to learn about researchers' interactions with IRBs in general and subsequently how they have dealt with data management and archiving

issues concerning the IRB in the past. Additionally, they were asked directly about their perception of the IRB's stance on data archiving.

### **3.2 Part B: Interviews**

The second component of the study was a series of short (10-15 minute) telephone interviews with members of a social and behavioral science institutional review board at a large research university, designed to explore their opinions and current knowledge concerning data archiving, including ethical concerns and how the intention to archive could affect the approval process.

Five out of thirteen Behavioral IRB members volunteered to be interviewed as a result of several recruitment emails and a visit by the researcher to a regularly scheduled IRB meeting. While the goal is similar to that of Part I, this population is small enough that members can be studied individually instead of surveyed en masse. The interview questions (see Appendix B) were designed to stimulate thought and discussion, with the intention of exploring, rather than explaining, members' attitudes. While a focus group could have allowed for more "pondering" of the issues among members (who already know each other and are used to interacting with one another), interviews reduce any disproportionate influence certain members might have had over others in a group setting, allowing opinions to be shared with the interviewer in confidence.

At the beginning of each interview, the IRB member was given an overview of the study's purpose to provide some context for the questions being asked, as well as some potential reasons for data archiving in social/behavioral science and a short description of what a digital repository does. The six questions were not necessarily

asked in order, but instead as discussion led to them. The first question asked for general thoughts or comments participants had after hearing the introduction and purpose of the study. Other questions asked about specific ethical issues that need to be addressed for archiving, and the IRB member could recall seeing the issue dealt with in applications. Each member was asked about their preferences regarding applications, including how and where data archiving might be addressed, and any affect the intention might have on their approval process. Finally, participants were asked directly to what extent they felt comfortable with de-identified social and behavioral science data being archived.

### **3.3 Part C: Content Analysis**

In the third part of this study, a sample of Informed Consent forms were examined to gain insight into the language typically used by researchers to describe their data management plans. The Informed Consent form, officially called "Consent to Participate in a Research Study (Adult Subjects)," is a place where researchers must explain succinctly, in lay terms, what will occur in the study, for the benefit of their subjects. This artifact was chosen for this study because here is where the most succinct "data destruction" language has been found and where we might gain best insight into the researcher's future plans for their data (if they have plans). While a waiver can be obtained under certain circumstances, most studies involving direct interaction (and data collection) from human subjects require a signed consent document. In addition, IRB members pay close attention to consent forms as they look to protect subjects' safety and confidentiality in their review process. Although content analysis offers straightforward results of what the artifacts contain, at the same time it offers little insight as to *why* the

result is occurring. Informed Consent forms provide evidence into language used by researchers to describe their data plans, they provide no explanation of why this language was used or how much deliberation was involved at all.

The Informed Consent forms examined were obtained, with permission, from Behavioral IRB records. All [94] Behavioral IRB applications on record for 2008 containing Informed Consent forms were included. The Informed Consent forms were examined for a statement or statements about what will happen to the data after the study is completed. Based on this information (or lack thereof), the forms were coded. Additionally, any interesting or exceptional language was noted.

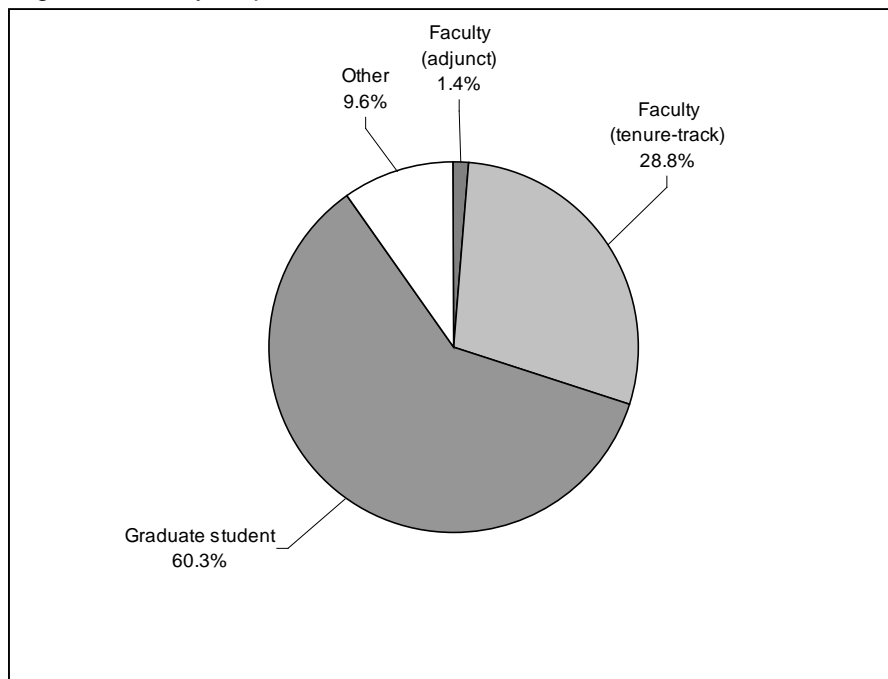


## 4. Results

### 4.1 Part A: Survey

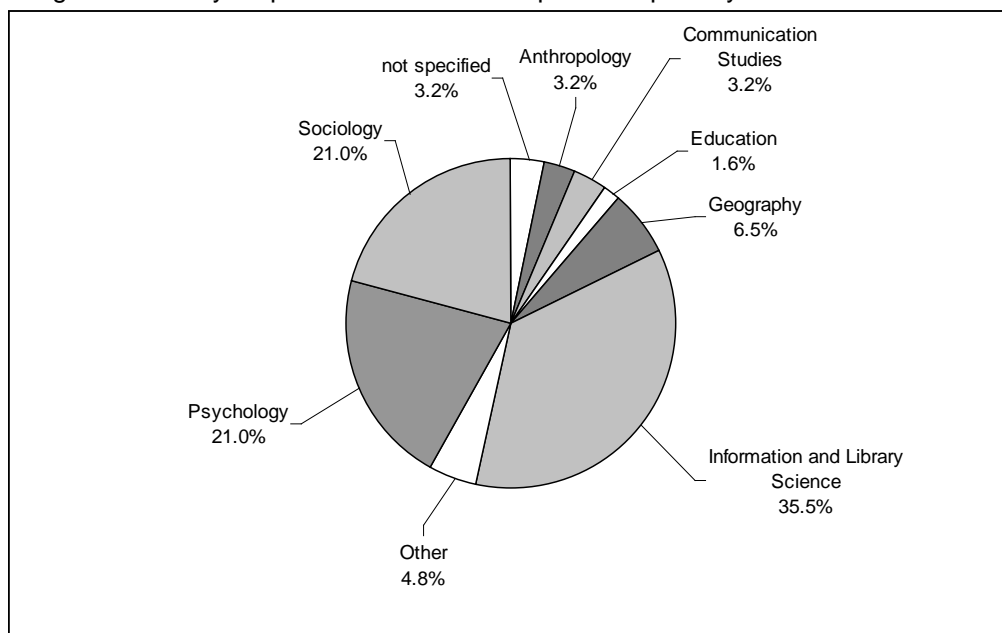
Approximately 700 emails were sent out, to as many social/behavioral science researchers possible on campus of a large research university, acknowledging that a significant percentage of these individuals may not study human subjects, therefore not requiring IRB approval (and subsequently not responding to the survey). There were 74 complete surveys returned, making the response rate just over 10%, although the response rate of total social/behavioral science researchers whose research requires IRB approval was likely higher. Of the respondents completing the survey, 29.9% were faculty members, 60.3% were graduate students, and 9.6% described themselves as "Other."

Figure 1: Survey respondents' academic status



These 74 respondents claimed seven different schools or departments as their primary affiliations, with 8% not responding or answering "Other."

Figure 2: Survey respondents' school or department primary affiliation



This distribution can be compared to a report provided by IRB staff, in which Psychology (102 applications) and Information & Library Science (82 applications) were the top two departments submitting applications to the Behavioral IRB in 2008. Education, Sociology, and Communication Studies were all in the top ten, with Anthropology and Geography close behind.

More than six out of seven (86.5%) of respondents had previously submitted IRB applications, further indicating that those who did not respond to the survey likely had not previously use the IRB or plan to in the future. This could provide explanation for the low response rate, although it is impossible to tell for sure.

Of the 64 respondents that had previously submitted IRB applications, 53 (82.8%) provided their data management plans from their previous applications. Respondents were asked to recall their answers to the following questions when answering:

A.4.12. **Data sharing.** With whom will *identifiable* data be shared outside the immediate research team? For each, explain confidentiality measures.

A.4.14. **Post-study disposition of identifiable data or human biological materials.** Describe your plans for disposition of data or human biological specimens that are identifiable in any way (directly or via indirect codes) once the study has ended. Describe your plan to destroy identifiers, if you will do so.

taken from study site's Application for IRB Approval of Human Subjects Research

Some respondents pasted their exact IRB application wording into the text box, while others simply described how they had answered this question. Their responses were coded on the following five-point scale<sup>2</sup>:

**Category 1:** Containing an explicit statement about data retention or archiving.

**Category 2:** Containing ambiguous language suggesting data retention.

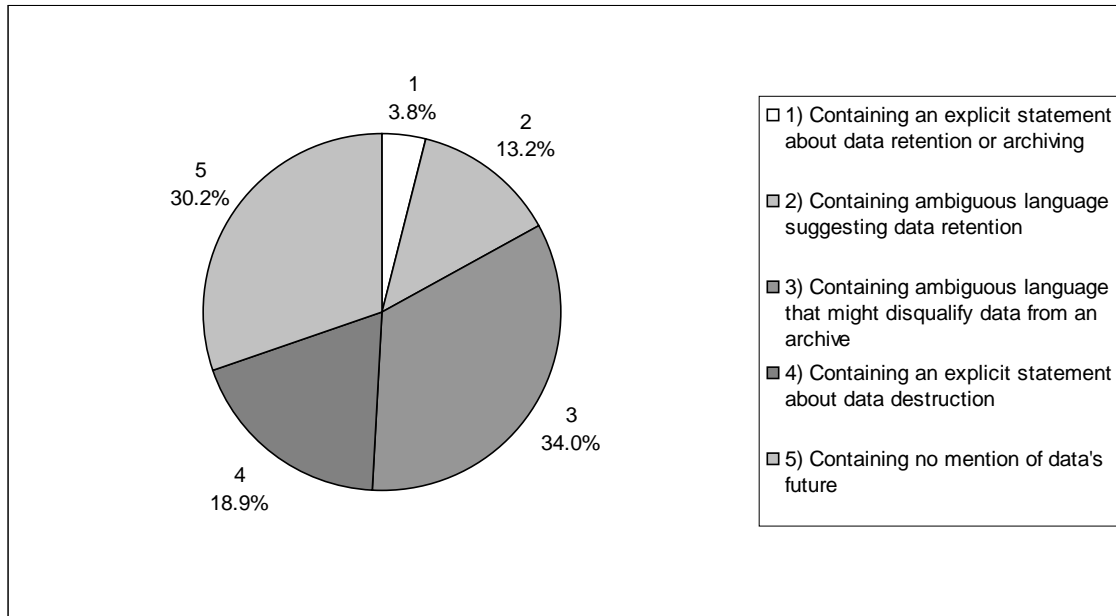
**Category 3:** Containing ambiguous language that might disqualify data from being archived. This includes suggestions that no data will be shared outside the research team, suggestion that data or some part of the data will be destroyed, or mention of retention/archiving that is for the sole use of the researchers.

**Category 4:** Containing an explicit statement about data destruction.

**Category 5:** Containing no mention of data's future.

The distribution appears below:

Figure 3: Distribution of coded "data management plans"



Some common ways researchers pledged to protect their subjects' confidentiality and anonymity included:

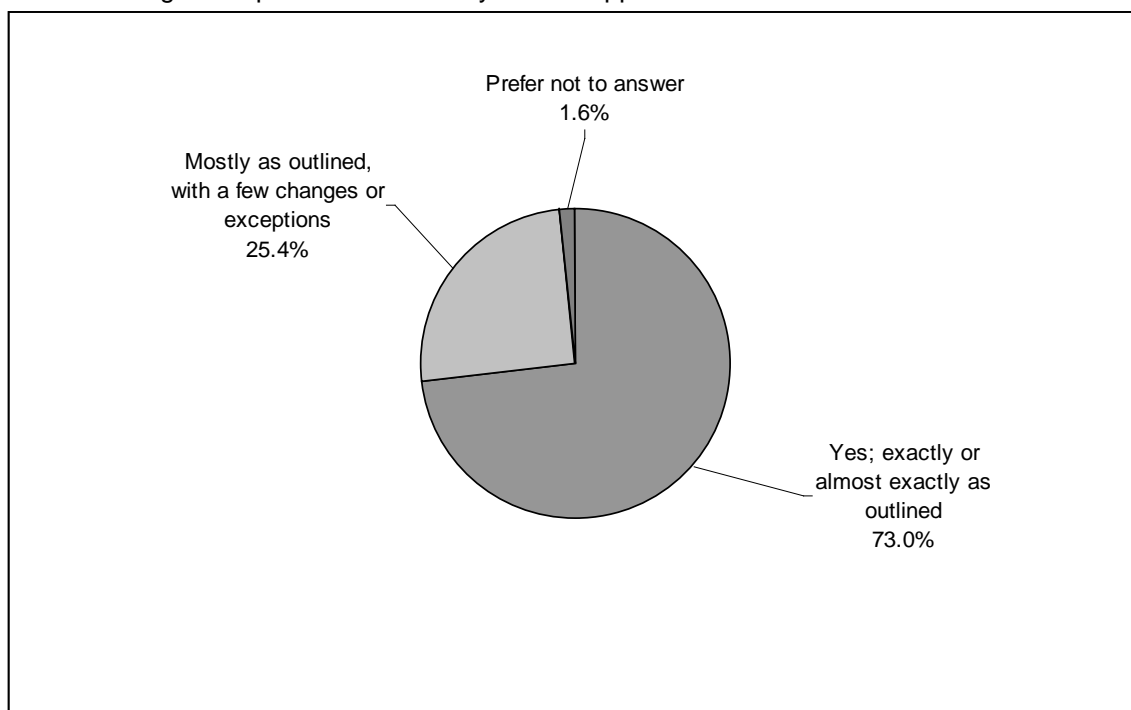
- Never collecting identifiers
- Using pseudonyms instead of names for subjects
- Using a carefully protected (and later destroyed) "linking document" that links subjects' identifying information to their pseudonym or code
- Destroying audio and/or video tapes after they had been transcribed
- Not sharing data outside the research team<sup>3</sup>
- Destroying data after the study was complete or after a specified period of time
- Maintaining data on a single computer only

Several researchers also mentioned professional associations' requirements, with statements like "the data will be retained for 7 years in keeping with APA requirements."

A few respondents declined to provide a data management plan because they only use secondary data in their research that has already been de-identified and made available, making the question irrelevant.

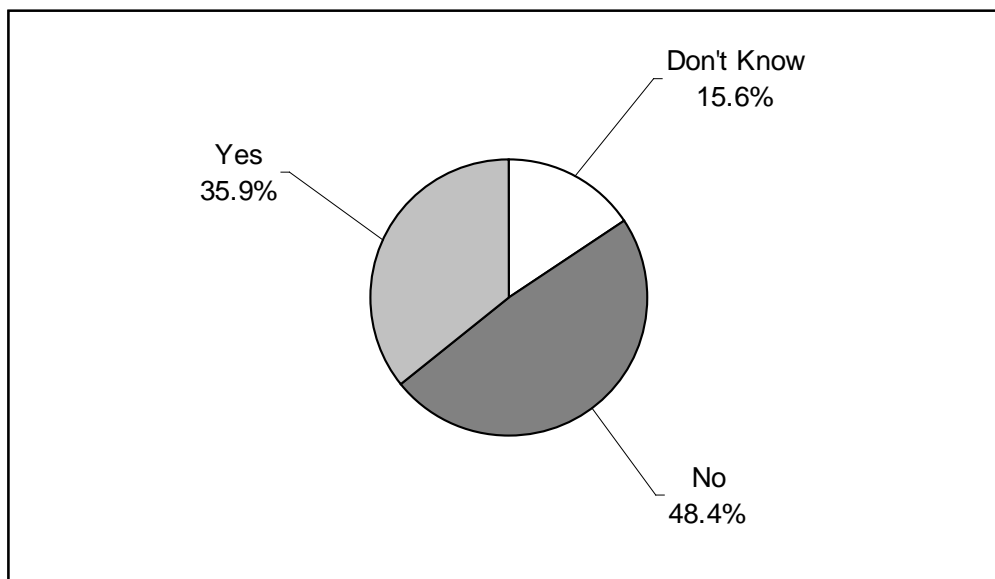
All respondents said that they followed their IRB data management plans exactly, almost exactly, or mostly with a few changes. No respondents chose the answer indicating that their data management practice had differed significantly from their outlined plan.

Figure 4: Answers to Question 5: "Did you follow, or are you following, your data management plan as outlined in your IRB application?"



Next, respondents were asked whether they would feel comfortable making their data available in a digital repository based on what they said in their IRB application. Approximately half (48.4%) said that they would *not* feel comfortable archiving their data based on their stated plans, while 35.9% would. 15.6% were not sure.

Figure 5: Answers to Question 6: "Based on your last application's wording, as it was approved by the IRB, would you feel comfortable archiving the de-identified data resulting from your study in a digital repository?"



Of those who said they were not comfortable (answered "No"), some common reasons cited were that their data is not de-identifiable or that the population is too sensitive, and that they were concerned about continued anonymity of their data. Others answered "No" because they had not included this plan in their original IRB application and/or had specifically stated that only the research team could access the data or had promised to destroy the data. A few said that their data was too context-specific and would not make sense or be of use to others. Several respondents were concerned that they had not informed their subjects about archiving and therefore did not feel they had received sufficient consent. One explained:

"My participants agreed to work with me, not with someone else. While they understand that the study will be a public document, I don't think my participants understand that their interviews and surveys could be used as data sets for other researchers. Doing that without having fully explained that to them already makes me very uncomfortable and is, I believe, unethical."

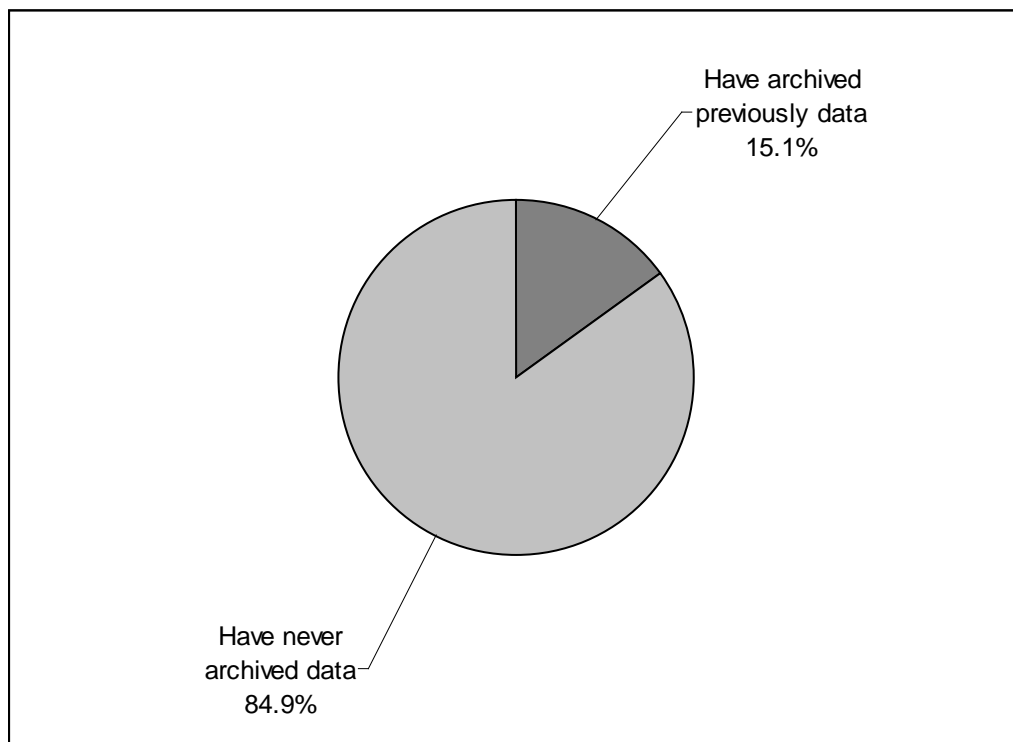
Those who said they would feel comfortable archiving their data (answered "Yes") also provided ethical reasons, the most common of which was that their data

included nothing personally identifiable and that deductive disclosure of identity or personal information was impossible or very low risk. A few reiterated their belief in open access to data, with one respondent stating "I do not agree to withhold information, only to make it non-identifiable." Some mentioned federal grant mandates. Another interesting response concerned ethnography research, with the respondent relating that his or her participants generally *want* their data to be identifiable and easily available to others. One respondent said he or she would feel comfortable archiving study data, but was "not sure anyone would want it."

The respondents that answered "Don't know" were generally uncertain about issues of ethics and control. Some felt their populations might be too vulnerable or that they might have needed to include data archiving in their consent documents but were not sure. One was not certain that he or she wanted others to be able to access their data without their control. Another said that his or her data would need to be de-identified, which would probably be too time consuming. One simply stated that he or she "[doesn't] know enough about this process."

All respondents, regardless of whether they had previously submitted an application to an IRB, were asked if they had ever archived datasets before. A majority (84.9%) said they had not, while 15.1% said they had.

Figure 6: Answers to Question 7: "Have you ever archived data from a study in a digital repository?"



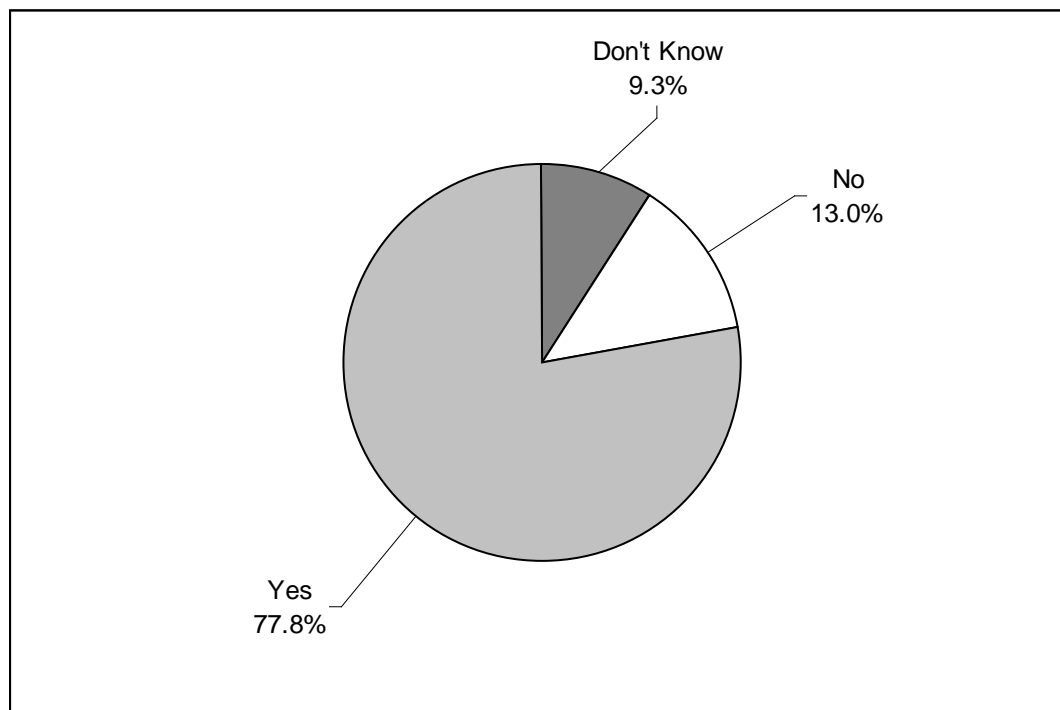
Of the 9 respondents who had previously archived study data:

- 4 (44%) explicitly stated in their IRB application that they intended to archive their data.
- 3 (33%) did not state in their IRB application whether they planned to save or destroy their data.
- 1 (11%) explicitly stated in the IRB application that the data would be destroyed.
- 1 respondent did not remember

Of the 62 who said they had never previously archived data, 77.8% said they would consider doing so. Thirteen percent said they would not consider archiving data and 9.3% did not know if they would consider it or not.



Figure 7: Answers to Question 8: "Would you consider doing so [archiving your study data in a digital repository]?"



Of those who answered "No," the most oft-cited reason was that they did not feel their data could "stand alone"; that it was context-specific and they would not want others manipulating their data without this background. It was also noted several times that the respondent would not consider archiving data when their subjects had not been informed about it; as one put it, "I would only consider [archiving my data] if I let participants know this might occur. I would not do it post-hoc." Another respondent who answered "No" noted that it might be "too much trouble." Interestingly, a respondent who answered "Yes" also stipulated that his or her answer would change to "No" if it required re-submitting to the IRB or took too much time.

Of those respondents who were unsure about archiving their study data (answered "Don't Know"), many mentioned ambiguity about who could access the data and what the

goals of the repository were. One stated that he or she could not think of a reason why one would archive data, but might be open to it if presented with a compelling reason.

Most of the respondents who were asked if they would consider archiving their study data chose "Yes," indicating that they *would* consider archiving their study data.

Many reasons were given, including some that have already been discussed in this study:

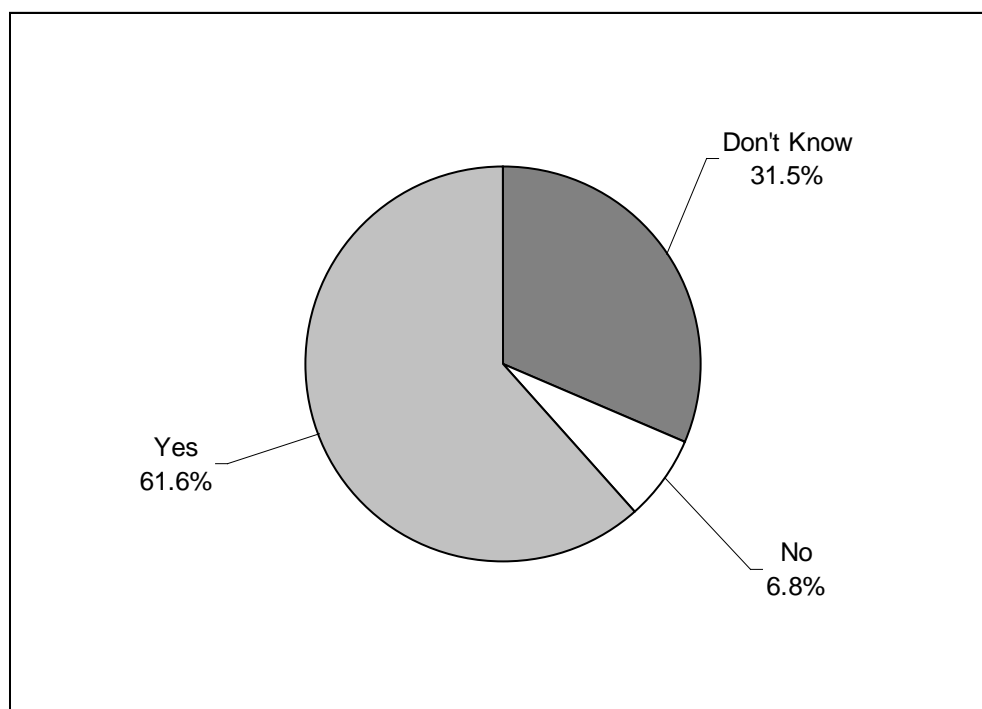
- It would be an effective way to share data with others.
- It would be useful for meta-analyses and comparative studies, considering that we have not always made full use of data collected.
- It seems more secure than storing paper copies.
- It increases academic transparency.
- The data might be helpful for others' research.
- It would allow for wider analysis and more minds trying to answer questions.
- It is resource-intensive to collect data and the knowledge gained could not easily be re-created.
- It supports the goals of advancing science.

One respondent felt he or she would be compelled to contribute because others have: "I've been helped by data from publicly available archives; if I collected original data I would feel obligated to add to this resource." Another also spoke to the advancement of science, musing that researchers have little excuse not to archive their data: "Why should the data not help science tomorrow, as much as science today? I'm surprised it's taken this long, actually. Mostly, the technology wasn't around to make it easy...now, we have little excuse."

A number of those who answered "Yes" did mention that they would only archive data if ethical concerns had been accounted for and if they had received IRB approval to do so. One emphasized the need for proper IRB language, admitting "I wouldn't know where to start if I was writing a repository into my IRB." Others talked about copyright and licensing issues and wanted to be sure of the security of the archive.

All respondents were asked whether they thought the IRB would approve an application that included data archiving/sharing plans.

Figure 8: Answers to Question 12: "Do you think the IRB would approve a study that included a plan to archive de-identified data for future use by you or other researchers?"



Only a few respondents answered "No," but those who did seemed to base their reasoning on past experiences with IRB applications. One respondent recounted:

I have received quite a few questions on every IRB I have submitted about why I do not plan to completely destroy the data from each study after completion, and it seems that they would like people to do so (based on the types of questions and concerns they have raised) as if to suggest that this is a bad thing to do.

Others cited protecting subjects as a possible reason that an IRB would not want data to be disseminated.

Respondents who answered "Don't Know" indicating that they were not sure about how the IRB would handle this situation, also expressed confusion and even disdain for the IRB process. Some thought it might depend on the wording in the application or the vulnerability of the population being studied. A very prevalent response from those who chose this answer was that the IRB and its processes are mysterious, causing them not even to hazard a guess. One respondent said, "I don't understand the reasoning behind the IRB--sometimes their review ignores big issues and picks on small ones. Thus, I can't really predict whether or not they would approve such a study." Also echoing the sentiments of others, another respondent recounted that "sometimes it is such a chore to get the IRB to accept a study that the tendency is to make it as easy as possible to get approval by promising to delete everything, even if one would prefer not." Still another hoped that the IRB would approve, but was reluctant to say because "in my experience the IRB has been very cautious (in some cases to the detriment of social science research)."

Nearly two thirds of respondents, however, thought that the IRB would approve an application that included intentions to archive data. The overwhelming feeling among these respondents was that as long as subjects' privacy and anonymity were properly protected, the IRB would have no problem with data archiving. Speculated concerns included de-identification and subjects' informed consent, as well as the security of the archive and specification measures the researcher would take to ensure these things. Many mentioned that their own use of archived data in their research and the current

existence of such data archives must mean that someone is approving of the practice. Some spoke from personal experience, relating that their IRBs had been accommodating of these requests in the past provided that the subjects were adequately protected. A few respondents also pointed out that in some cases, the IRB *must* approve--particularly because the National Institutes of Health (NIH) and other major granting agencies (especially publicly funded ones) are beginning to require that resulting data be made available for other researchers.

It is also worth noting that many responses included statements about the IRB's reason for existence and main goals--to protect subjects and uphold ethical standards--as a reason why archiving de-identified data would be approved. One respondent articulated this sentiment about the IRB's intentions: "I think as long as we are upfront with our participants of all possibilities of use (use of the responses), they'd be pretty much ok with anything." Along the same lines, several respondents acknowledged that IRB members are also interested in the advancement of knowledge and research: "They're scientists too...I'd be surprised if they didn't think it was valuable to [archive data]."

#### **4.2 Part B: Interviews**

Five members of a social/behavioral science institutional review board at a large research university were interviewed over the telephone; this sample represented approximately one third of all members of this board. The board consists of faculty members as well as community members, many of whom are advocates for certain populations. The interviews included representatives from both of these groups. The

results presented here are gathered from the interviewer's notes and the audio recordings that were used to verify direct quotations.

When asked about their general thoughts and feelings regarding data archiving in the social/behavioral sciences, all of the members interviewed responded positively, though several had reservations. One member was resolute that it was "not a problem," a sentiment also echoed by the other members interviewed. Another member came out in full support of data archiving, citing verification of results as a reason, but also pointing out that "a lot of time and effort goes into gathering the info and that it's great to use technology to get additional value out of it." Two members stated that they had no issues with researchers wanting to include data in archives, but were careful to say that everything must be properly de-identified first. One put this in context of the duties of an IRB member--to protect subjects' anonymity and confidentiality as much as possible and protect them from harm or embarrassment--assuring that "as long as we are able to take care of those areas, I'm fine with it." Another member had "never really thought about it," but supposed he or she was in favor of it, if it was going to advance the sciences. This member also confessed that, in all honesty, he/she did not really care one way or the other about the issue.

Most IRB members interviewed mentioned "anonymity," "confidentiality," and "de-identification" at some point in the discussion as the important ethical issues regarding data archiving in the social sciences. Another ethical gray area one member discussed was a current practice where some data *are* kept ("archived") for a certain amount of time--often almost a decade--because of professional associations (such as the American Psychological Association) or journals mandating that they do so. The

question, then, is whether it would be ethical to extend that period of time indefinitely, or how one might go about getting consent for that.

None of the members could recall seeing applications in the past that addressed archiving study data in a repository of some sort, although one recounted a few instances where biological samples were to be retained. Although one member had not seen any IRB applications that discussed data archiving, he or she had seen instances in federal grant applications, where data sharing is sometimes mandated if the budget is over a certain amount.<sup>4</sup>

Four of the five members interviewed said that they would prefer to see data archiving plans explicitly indicated on IRB applications, as opposed to making do with ambiguous language. One member was very adamant that ambiguous language is often used for purposes of deception and that "there should never be something done through deception for the benefit of the project itself." This member was against ambiguous language in the entire application so that communication between subject and researcher can be as clear as possible. One member thought that ambiguous language has been the default but because archiving is becoming more prevalent and easier, "from a legal standpoint and an informed consent standpoint it would be preferable to be clear about it upfront." Another member was not comfortable with just anyone being able to archive data after the fact, so felt that the researcher should have the foresight to think of future potential uses and make a case for them in their initial applications. On the other hand, one member had no problem with it being generally known and accepted that de-identified study data may be archived even if the application does not specifically say so.

As for where on the application the IRB members would like to see data archiving addressed, three discussed including a straightforward statement on the consent document for the subject. One member suggested that there be standard "boilerplate" language created for the consent document saying that a de-identified version of the data collected will be available for future analyses, and that this could be added into the existing section that discusses protection of confidentiality and uses of the data. Another IRB member imagined an informed consent form that allowed the subjects to choose whether their data was included in archiving or not. That way they would have the option to consent to the current study and then decide separately whether they also consent for their data to be used in future studies--they would have the choice whether to opt in or opt out.

Several also discussed researchers' justifying data archiving within the IRB application itself. One member found it important for the researcher to justify the reason for archiving data in his or her IRB application, including what increased benefits archiving would bring forth. The same member pointed out that archiving for unforeseen future uses was not a good justification, because this could be argued for anything. Another IRB member suggested creating a new section on the IRB application that specifically dealt with data archiving, asking the researcher to declare this intention and then state how the data will be stored, that it will be completely de-identified, and how they are going to safeguard the anonymity of their subjects.

When asked how a data archiving plan might affect the approval process, all of the IRB members interviewed essentially agreed with one member's statement that it would be "just one more thing to consider." Several members reiterated that the subjects must be protected, but none thought it would adversely affect an application's approval. One



mentioned that something truly not de-identifiable would give pause, but barring this exception, "if anything, [seeing a well-crafted data archiving plan in an application] would give me a warm fuzzy feeling--make me feel better about it--right on, doing the right thing!"

Finally, each of the five IRB members interviewed was asked "To what extent are you comfortable with social science data being archived for future use?" Their replies are as follows, in random order:

- I am very cautious.
- I'm comfortable with it.
- I guess I'm totally comfortable.
- I like the idea a lot. Enthusiastic proponent.
- I think I'm very comfortable with it. Like I said, as long as the subject's identity is protected is for me I think the biggest issue.

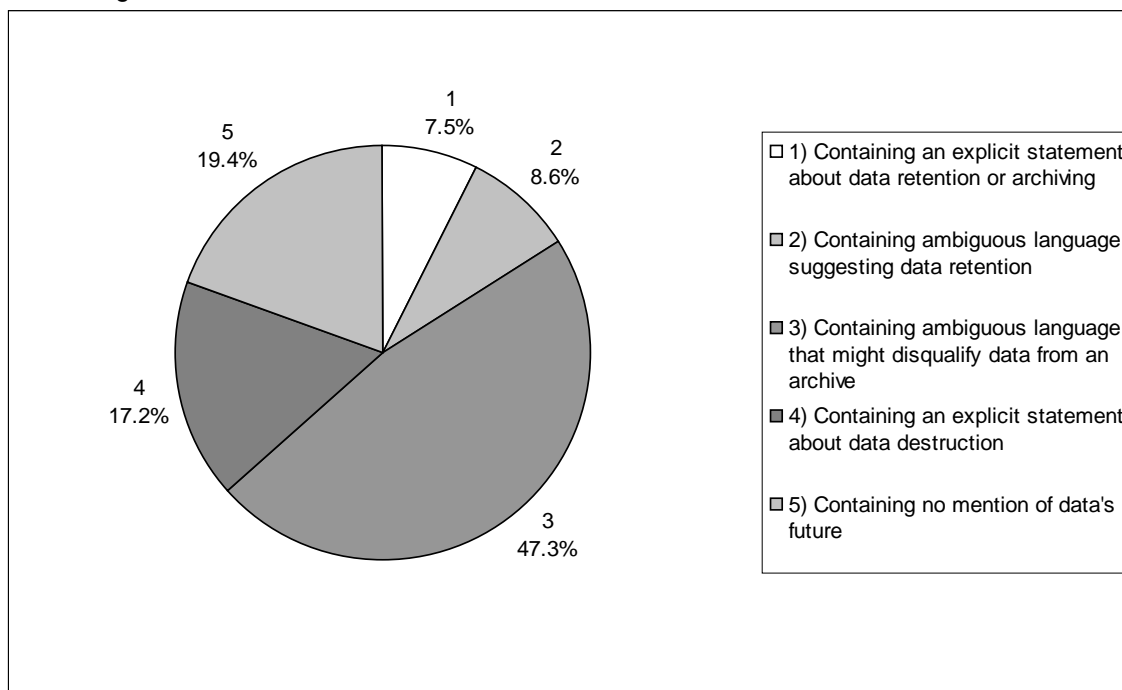
When asked if there was anything else they would like to say on the issue of data archiving in social and behavioral science, the only comments were from two members who both noted that this topic is an important and timely one that they were glad to begin considering. One member concluded by stating, "I believe that [we are going to receive] more and more of those kinds of requests and we need to be prepared to handle them."

#### **4.3 Part C: Content Analysis**

Ninety-four blank Informed Consent forms from studies submitted to the Behavioral IRB in 2008 were analyzed, looking for statements about the future of the data

collected. These were coded using the same scale as the survey respondents' data management plans (found in Figure 3). The findings are presented in Figure 9, below, along with examples of language found in each category following.

Figure 9: Distribution of coded blank Informed Consent forms



**Category 1:** Containing an explicit statement about data retention or archiving.

Examples:

- "Survey data will be compiled into a database and the data may be used in future publications or presentations."
- "The information will be maintained for at least 20 years for possible use in additional studies."
- Data can be destroyed after transcription per subject wishes, otherwise maintained safely "in case it proves to be useful in future anthropological investigations."
- Consent form seeks permission to release video for possible educational/scientific uses.

**Category 2:** Containing ambiguous language suggesting data retention.

Examples:

- "The interviews will be kept for an unspecified amount of time"
- All survey data kept on personal computer and will not be destroyed.

**Category 3:** Containing ambiguous language that might disqualify data from an archive. This includes suggestions that data will not be shared outside the research team, suggestion that data or some part of the data will be destroyed, or mention of retention/archiving that is for the sole use of the researchers.

Examples:

- "All video and audio data will be deleted and/or destroyed before the student participants' 18th birthday." [sic]
- "All records stored and permanently archived in our locked lab at UNC"; only research team will have access.
- "The transcriptions will be kept by the PI for future use and all audio tapes will be destroyed."
- The audiotapes will be destroyed within 6 months of the study's completion.

**Category 4:** Containing an explicit statement about data destruction.

Examples:

- "After data are analyzed, we will destroy all information collected."
- "The raw data will be kept for 7 years, keeping with requirements of academic journals, after which it may be destroyed."

**Category 5:** Containing no mention of data's future.

It is important to note that although the same set of codes was used to analyze the blank Informed Consent forms and researchers' responses to the survey question about their data disposition plans, these two sets of data cannot be meaningfully compared statistically, due to their differences. Informed Consent forms are only included in studies employing certain methods such as interviews and experiments, while the survey

respondents could be referring to any number of study methods that involve human subjects. The two inquiries were meant to be compared qualitatively only.

## 5. Discussion

The findings of this three part study suggest a number of things about social and behavioral researchers and the IRBs that serve them

- Language used on IRB applications (as well as language that *is not* used), and the underlying ethics of the issue are indeed a barrier to data archiving in social science.
- Neither researchers nor IRB members give significant thought to data archiving in early stages of social/behavioral science research.
- Social/behavioral researchers and IRB members are both interested in protecting subjects and upholding professional ethics.
- There is evidence that social/behavioral IRBs have and will approve applications that include data archiving plans.
- Data archiving in social and behavioral sciences is becoming a bigger issue that IRBs and researchers will need to consider seriously.

Two parts of this study examined the language researchers use on their IRB applications and in the consent documents provided for their subjects. Both their stated data plans (Figure 3) and blank Informed Consent forms (Figure 9) showed similar patterns: a very small percentage included explicit language describing archiving of data and a majority displayed language that might later cause data archiving to be unethical.

Although there were many instances where researchers explicitly stated that they would destroy all of their data after analysis or after the study was complete, interestingly it was much more common to see other ambiguous language surrounding the data management plan. The "Category 3" code was described as "Containing ambiguous language that might disqualify data from an archive" in order to capture these various instances. There were a few different ways researchers' language could be included in this category. The specific IRB application questions they were to consider asks with whom *identifiable* data will be shared and how *identifiable* data will be disposed of once the study has ended<sup>5</sup>. Researchers' statements made in response to these questions often dropped the word "identifiable"--so that they were claiming not to share any data outside the research team, or that all data will be destroyed after the study has ended. Intentional or not, these statements made the researchers' intentions for their data unclear and their data unusable by others. When this same issue was observed on blank Informed Consent forms, even if the researcher *meant* identifiable data, this language could have been misleading to the subjects. Another instance included in Category 3 was when the researcher stated that part of their data would be destroyed, whether it was audio and video tapes, documents linking identifiers to codes, or survey forms that had been recorded digitally. Although this is common practice, talk of destruction alone could be confusing or misleading, especially to subjects. Some very clearly stated that the data would be retained or archived, but that it would be solely for the researchers' own future use. Others mentioned that only the researchers or PI would be able to view the data. None of these promises outright destruction of data, but still makes archiving ethically vague. The blank Informed Consent forms showed similar signs of indistinctness.

The comments from IRB members, however, indicate that these distinctions are somewhat futile, because the members interviewed would prefer to see explicit mention of data archiving plans. To most IRB members interviewed, even suggesting that datasets might not be destroyed (as in Category 2) is not adequate or considered the most ethical path.

It is not surprising that the language used in IRB applications about data archiving was less than clear. Perhaps the IRB member interviewed summed it up best when he/she said archiving of social science data was something he/she had "never thought about before." Although many members of both populations studied had opinions on the issue, the data gathered suggests that researchers are not thinking about the future of their study data at the time of the IRB application. In addition to the large amount of ambiguous language and possibly unintentional declarations already discussed, approximately 30% (Figure 3) of researchers' stated data management plans and almost 20% of blank Informed Consent forms (Figure 9) contained no mention of the study data's future at all. More than three fourths (77.8%) of those who had not previously archived data said they would consider doing so. However, nearly half (48.4%) of researchers who responded to the survey said that they were not comfortable archiving their data based on how they had presented the study to the IRB (Figure 5) suggesting that those who would consider it were probably not considering it at the time they wrote their IRB application. In light of the IRB members' interview responses obtained in this study, this lack of foresight for data archiving is problematic. Since IRB members interviewed almost exclusively wanted to see researchers address this directly in their applications (see Section 4.2)

failing to mention data archiving plans--especially to one's subjects--might permanently eliminate the option to archive the study data for others' use.

There is also evidence that researchers and IRB members are both ultimately interested in upholding the same ethical principles, primarily that of protecting subjects--something that may be obvious but is worth pointing out as a reminder. The IRB application is a plan for a study that upholds ethics agreed upon by both the IRB and the researcher. It may seem like researchers are "going up against" the IRB when really this is not the purpose for the board's existence. Data from this study support the idea that IRB members and social science researchers are essentially interested in the same goal: to protect subjects from unnecessary harm.

It is clear that ethics and IRB approval are important to the social/behavioral science researchers who completed the survey. Respondents repeatedly cited IRB approval and protection of subjects as reasons for doing or not doing things--for instance, when asked "Based on your last application's wording, as it was approved by the IRB, would you feel comfortable archiving the de-identified data resulting from your study in a digital repository?" (Figure 5) justifications for all answers ("Yes," "No," and "Don't Know") included assessing how identifiable the data is and how sensitive the population studied. Of those who had previously archived study data, four out of nine informed the IRB that they did this (or were going to do this). Although some researchers expressed frustration and bafflement about the IRB's process, many more indicated understanding of the board's purpose and position and 61.6% were confident that the board would not object or stand in the way of data archiving as long as it was done properly.



Indeed, this study's findings suggest that this is correct: the content analysis revealed that the IRB has approved studies with data archiving plans in the past (Section 4.3) and interviews revealed IRB members ranging from cautiously supportive to outright enthusiastic about the practice of data archiving (Section 4.2). Additionally, some survey respondents referred to their own experiences where the IRB has approved archiving of study data. The IRB members were careful to specify in the interviews that they were still cognizant of subjects' anonymity and confidentiality, as well as their right to refuse. One member suggested development of standardized language to inform subjects of data archiving plans, which seems necessary: a survey respondent echoed the sentiments of others, admitting "I wouldn't know where to start if I was writing a repository into my IRB." In general, IRB members interviewed agreed that archiving of data was not something that should be done post-hoc and must instead be specifically addressed early on. Several members also said that some justification would need to be made in the application: archiving for its own sake would not be acceptable. With these concerns accounted for, all of the IRB members interviewed agreed that a stated plan to archive de-identified data would not affect the application process, including the speed of approval. Of course these are just the opinions and ideas of a few IRB members, but give us good insight into how others might feel as well.

Although it is certainly not a new practice, data archiving in social/behavioral science is becoming easier and more common, meaning that IRBs and researchers alike should begin to learn about it and consider it seriously if they have not already. Both groups mentioned requirements of professional associations like the APA, and funding agencies like the NIH that have stipulated data retention; NIH funding also requires

providing public access to data collected. Researchers are well aware that others are archiving datasets and many mentioned making heavy use of these in their own work. Additionally, researchers and IRB members seem largely cognizant of the benefits of archiving, and a large percentage (77.8%) of those surveyed who had never archived data said they would consider doing so. If researchers and IRB members are not thinking about social science data archiving already, they will need to start doing so in the near future.

This study's findings should also be of interest to archives and repositories looking to collect researchers' data. Various researchers expressed that they want archiving to be easy, that they would need good mechanisms for providing context for their data, and that they need to be presented with compelling reasons to contribute to a repository. Members from both groups mentioned concerns about the repository's security and the archive's mission: both endeavors that are heavily discussed within the information science community but that need to be shared with the repository's users as well. Finally, there is evidence that social science researchers are using data archives and repositories in their work, indicating the ongoing importance of storing and managing data into perpetuity.

## **5.1 Limitations**

It is also important to note the limitations of this study. For one thing, the response rate in the survey was relatively low--around 10%--and due to the survey's structure, some survey items had a lower number of respondents. Information and Library Science faculty and graduate students are well represented in the sample, and although

this is appropriate due to the large number of applications submitted to the IRB by this group, it is possible that they would be more aware of digital repositories or more supportive of them than other social science researchers, skewing results. Finally, the three parts of the study were carried out at a single large research university. While it is likely that this study's findings can be considered representative of similar universities, every IRB is different, as are different populations of researchers. Many of this study's recommendations for IRBs and researchers, however, hold true regardless of an institution's particulars.

## **5.2 Recommendations**

- **Early Intervention by Repositories**

IRB members and researchers alike seem to agree that the ethical issue involved in archiving social science data is not only notifying the IRB of this intention, but notifying one's subjects. While IRB applications can be modified later with a form or two, getting subjects' consent after the fact is much more difficult, if not impossible. Therefore, it is critical that researchers think about archiving data while a study is still in the planning stage. Archivists and those recruiting collections for digital repositories must also consider intervening early in the process in order to assist researchers in taking the proper steps to create archivable data.

- **Discussion within Institutional Review Boards on Data Archiving in Social/Behavioral Science**

As one IRB member interviewed pointed out, boards can expect to see archiving of data in social and behavioral sciences more in the future and should be prepared to

handle these requests. Only a few members of a single board were questioned in this study, and although their responses give us good insight into what a Behavioral IRB might be looking for, every board will have to decide this for itself. Discussing the issue will encourage consistency among members, making their jobs easier in addition to benefiting researchers submitting applications.

- **Guidance Provided by IRBs and Repositories**

By the same token, institutional review boards that have reached consensus on how data archiving should be approached would do well to provide guidance for its applicants. This could be in the form of sample "boilerplate" language for consent documents or guidelines about ethical ways to approach data archiving. If IRB applications are amended to include a section on data archiving, as one IRB member suggested, the board might provide examples of what they would consider good data management behavior. Additionally, repositories themselves could provide language for researchers to include in their applications, or to IRBs directly, that discusses the archive's procedures, terms of use, and security. Including this kind of guidance will not only make it easier for researchers to navigate the IRB process, but will also act as a publicity tool for the repositories trying to encourage deposit and use.

- **Continued Education**

It is clear from researchers' survey responses in this study that they are well versed in good research practices regarding protection of their subjects and concepts of sensitive populations and deductive disclosure. To advance the practice of data archiving in the social and behavioral sciences and to satisfy the IRB that subjects are being kept from harm, it is important to continue to uphold these values at all levels of education, as

the APA Presidential Task Force (2007) also emphasized. Additionally, education on de-identification of data, especially where it is not always straightforward (e.g. interviews) would be useful, as would education on best practices for creating data to be archived-- which archivists and librarians would be apt to share.

One final comment regarding education is that IRBs and researchers alike are served by increased transparency from both sides. IRBs ask for clear and concise language in applications so that they know exactly what the researcher plans to do so that they can best ensure that subjects are protected. Researchers want to know what is acceptable to an IRB so that they can design and describe a study and receive approval as quickly as possible. At the end of the day, both groups want to uphold professional ethics and protect subjects from unnecessary harm. When an issue like data archiving, that has possible ethical implications, comes to light, debate should occur and it should occur out in the open. Continuing research and other scholarly communication regarding IRBs and the communities they serve will help the IRB and researchers constantly improve and update the process, enabling the advancement of science.

## 6. Notes

1. For examples of definitions of institutional repositories, see Lynch (2003) and Crow (2002).
2. These same codes were used to analyze the Informed Consent forms in Section 4.3.
3. Although the IRB application section asks how identifiable data will be protected, many replies include statements that drop this word, instead saying things like "No data will be shared outside the research team" or "All data will be destroyed within 6 months of the study's completion." A few did specify that they were referring only to identifiable data being protected or destroyed.
4. Since none could recall applications with data archiving plans, the question "Was the researchers' handling of it in the application satisfactory?" was not asked to any member.
5. Asking how identifiable data is to be destroyed is very different than asking how identifiable (or any) data will be retained. The language of the question implies that destruction is necessary.

## References

- Allard, S., Mack, T. R., & Feltner-Reichert, M. (2005). The librarian's role in institutional repositories: A content analysis of the literature. *Reference Services Review, 33*(3), 325-336.
- APA Presidential Task Force on Institutional Review Boards and Psychological Science. (2007). *Recommendations of the 2007 APA Presidential Task Force on Institutional Review Boards and Psychological Science*. Retrieved October 20, 2008, from [http://www.apa.org/science/rcr/IRB\\_Report2007.pdf](http://www.apa.org/science/rcr/IRB_Report2007.pdf)
- Ashcraft, M. H., & Krause, J. A. (2007). Social and behavioral researchers' experiences with their IRBs. *Ethics & Behavior, 17*(1), 1-17.
- The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. (1978). DHEW Publication No. (OS) 78-0012.
- Candilis, P. J., Lidz, C. W., & Arnold, R. M. (2006). The need to understand IRB deliberations. *IRB: Ethics & Human Research, 28*(1), 1-5.
- Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. <[http://www.arl.org/sparc/bm~doc/ir\\_final\\_release\\_102.pdf](http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf)>.
- Davis, P. M., & Connolly, M. J. L. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine, 13*(3-4).
- Easter, M. M., Davis, A. M., & Henderson, G. E. (2004). Confidentiality: More than a linkage file and a locked drawer. *IRB: Ethics & Human Research, 26*(2), 13-17.
- Finifter, B. M. (1975). Replication and extension of social research through secondary analysis. *Social Science Information/Information Sur Les Sciences Sociales, 14*(2), 119-153.
- Fitch, K. L. (2005). Difficult interactions between IRBs and investigators: Applications and solutions. *Journal of Applied Communication Research, 33*(3), 269-276.
- Foster, N. F., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine, 11*(1).

- Gierveld, H. (2006). Considering a marketing and communications approach for an institutional repository. *Ariadne*, 49.
- Green, A. G., & Gutmann, M. P. (2007). Building partnerships among social science researchers, institution-based repositories and domain specific data archives. *OCLC Systems and Services*, 23(1), 35-53.
- Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. *Journal of Digital Information*, 8(2).
- Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy*, 3(2), 327-336.
- Markey, K., Rieh, S. Y., St. Jean, B., Kim, J. & Yakel, E. (2007). Census of institutional repositories in the United States: MIRACLE Project research findings. Washington, D.C.: Council on Library and Information Resources. CLIR Publication No.140, 167 p. <http://www.clir.org/pubs/reports/pub140/contents.html>
- McDowell, C. S. (2007). Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, part 2. *D-Lib Magazine*, 13(9-10).
- Parry, O., & Mauthner, N. S. (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology*, 38(1), 139-152.
- Scott, C. R. (2005). Anonymity in applied communication research: Tensions between IRBs, researchers, and human subjects. *Journal of Applied Communication Research*, 33(3), 242-257.
- Watson, S. (2007). Authors' attitudes to, and awareness and use of, a university institutional repository. *Serials*, 20(3), 225-230.



## Appendix A: Survey Questions

1.) What is your status at UNC?

- Faculty (tenure-track)
- Faculty (adjunct)
- Graduate student
- Staff
- Other (please describe your status):

2.) What is your primary departmental affiliation?

- Anthropology
- City and Regional Planning
- Communication Studies
- Education
- Geography
- Information and Library Science
- Psychology
- Social Work
- Sociology
- Other

3.) Where have you submitted IRB applications? (check all that apply)

- At UNC-Chapel Hill
- At another university
- At another institution such as a hospital or research center
- Other (please describe):
- I have never submitted an IRB application and/or am currently in the process of submitting my first one

To answer the following question, think back to your answers for the last IRB application you submitted.

4.) Briefly describe your plan for managing subjects' data (digital, paper, passwords, de-identification, etc.) Think about your answers to the following questions (taken from UNC-CH's Behavioral IRB Application), or refer to your application if you'd like.

A.4.12. Data sharing. With whom will identifiable data be shared outside the immediate research team? For each, explain confidentiality measures.

A.4.14. Post-study disposition of identifiable data or human biological materials. Describe your plans for disposition of data or human biological specimens that are

identifiable in any way (directly or via indirect codes) once the study has ended. Describe your plan to destroy identifiers, if you will do so.  
taken from UNC-CH's Application for IRB Approval of Human Subjects Research

5.) Did you follow, or are you following, your data management plan (described in the previous question) as outlined in your IRB application?

- Yes; exactly or almost exactly as outlined
- Mostly as outlined, with a few changes or exceptions
- No; data management practice has differed significantly from my outlined plan
- Prefer not to answer

6.) Disciplinary data repositories such as the Odum Institute here at UNC or the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan, and institutional repositories, such as the one currently under development at UNC, store and make study data available over time according to data use agreements.

Based on your last application's wording, as it was approved by the IRB, would you feel comfortable archiving the de-identified data resulting from your study in a digital repository? Please answer YES or NO and provide some explanation for your response.

- Yes
- No
- Don't Know

Why or why not?

7.) Have you ever archived data from a study in a digital repository?

- Yes (skip to 11)
- No (continue on to 9)

9.) Would you consider doing so?

- Yes
- No
- Don't know

10.) Why or why not?

(skip to 12)

11.) Did you alert the IRB that you archived your study data?

- Yes; I explicitly stated in my IRB application that I would archive my data
- Yes; I submitted a modification form to alert the IRB of my plans
- No; I didn't state in my IRB application whether I would save or destroy my data
- No; I explicitly stated in my IRB application that I would destroy my data
- Don't remember
- Other (please elaborate):

12.) Do you think the IRB would approve a study that included a plan to archive de-identified data for future use by you or other researchers?

- Yes

- No

- Don't know

Please explain your answer:

## **Appendix B: Interview Instrument**

Thank you for agreeing to speak to me today. Let me remind you that you may elect not to answer any question, for any reason, at any point during this discussion.

What I'd like to discuss with you over the next 15-20 minutes are your thoughts and feelings, as a member of an Institutional Review Board, on study data generated by social and behavioral science researchers--properly de-identified of course--being archived in digital data repositories for future use.

A digital repository stores and makes digital documents available over time, for future use by the original data collectors, other researchers, or even the general public, according to use agreements. Archiving of digital data sets is becoming more and more common in the sciences and social sciences for reasons such as verification or replication of results, or completely new uses of data that has already been collected. In social and behavioral science, there are disciplinary repositories, such as the Odum Institute at UNC or ICPSR at the University of Michigan, and institutional repositories, such as the one currently under development at UNC, that will preserve digital files and information about them so that they are easily searchable and understandable, and so that technological changes don't render them inaccessible.

What I am interested in today are any ethical concerns you might have regarding this practice, your impressions of this practice as it is today, how you'd like to see this issue addressed by researchers in their IRB applications, and how this expressed intention might impact your granting approval. Before we begin, do you have any questions or need any clarification on the concepts of digital archives, institutional repositories, or data archiving?

- I'll start out by asking you to talk about any general thoughts or feelings you have about behavioral science data being archived to be reused.
- What are the ethical concerns that need to be addressed for you to approve of this practice?
- Have you seen applications that address data archiving?
- Was the researchers' handling of it in the application satisfactory? (Why or why not? How might he or she change this?)
- Would you prefer that a researcher explicitly indicate his or her intentions to archive their data in their application or is ambiguous language okay? (For instance, would it be

acceptable to simply omit mention of data destruction or would you rather see the researcher say that he or she intends to make it available?)

- On what pieces or questions of the application would this need to be addressed?
- How would a researcher's intention to archive study data affect your approval process? (Do you think it would affect the speed of the process?)
- So now that we've talked about these issues, I want to pose the question more directly: to what extent are you comfortable with social science data being archived for future use?

We've now reached the end of my questions: is there anything else you'd like to say on this topic that we haven't already covered?