

Sophia N. Lafferty-Hess. Sharing Research Data in the Social Sciences: The Role of Journal Policies. A Master's Paper for the M.S. in I.S. degree. April, 2014. 51 pages.
Advisor: Christopher A. Lee

This paper explores the relationship between social science journal data sharing policies and online data availability. The study was conducted to understand the effectiveness of data sharing policies on influencing researchers to share their data.

Data availability for 65 articles published in social science journals with three types of data sharing policies (mandatory, recommended, or no policy) were examined. The study also examined the type and location of the data. A low percentage of researchers were found to share their original research data and the majority who shared their data were bound by a mandatory data replication policy. These findings suggest that there is a relationship between strong data sharing policies and the rate of data sharing. Journals can use these findings to inform decisions on maintaining, expanding, or enforcing policies.

Headings:

Data libraries

Data sharing

Social science archives

Periodicals

Information retrieval

Institutional repositories

SHARING RESEARCH DATA IN THE SOCIAL SCIENCES:
THE ROLE OF JOURNAL POLICIES

by
Sophia N. Lafferty-Hess

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
Information Science.

Chapel Hill, North Carolina

April 2014

Approved by

Christopher A. Lee

TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
LIST OF TABLES AND FIGURES.....	2
CHAPTER	
I. INTRODUCTION.....	3
II. LITERATURE REVIEW.....	6
III. METHODS.....	15
IV. RESULTS.....	18
V. DISCUSSION & CONCLUSION.....	25
REFERENCES.....	35
APPENDIX A: DATA AVAILABILITY SEARCH STRATEGY.....	42
APPENDIX B: JOURNALS AND DATA SHARING POLICIES.....	43
APPENDIX C: TOP 100 SOCIAL SCIENCE JOURNALS.....	47

LIST OF TABLES AND FIGURES

Table 1. Journal Data Sharing Policy Examples.....	19
Table 2. Distribution of Data Sharing Policies	20
Table 3. Existing vs. Original Data.....	20
Table 4. Data Types of 65 Articles	21
Table 5. Data Availability.....	21
Table 6. Original Data Availability by Policy and Location	22
Table 7. Data Types for Available Existing and Original Data	24
Figure 1. DCC Curation Lifecycle Model	9
Figure 2. Data Included as Additional Materials on Journal Website	22

CHAPTER ONE

INTRODUCTION

Today, we are experiencing what has been called a “data deluge” (Hey & Trefethen, 2003). One kind of data that has significant intellectual and public value is research data, which has been defined as “recorded factual material commonly accepted in the scientific community as necessary to validate research findings” (OMB, 1999, para. 36). Although there has also been a consistent push by various stakeholders, including scholars, funders, professional organizations, and journal editors to encourage data sharing, research data is not generally being preserved or made accessible at the rate many would hope.

Over the past decade, external stakeholders have been encouraging researchers to share their data largely through the implementation of policies. In 2003, the National Institutes of Health (NIH) implemented their data sharing policy, which states that NIH “expects and supports the timely release and sharing of final research data” (NIH, 2003, para. 2). In 2011, the National Science Foundation (NSF) introduced data management plan requirements for all grant proposals (NSF, 2011). In February 2013, an Office of Science and Technology Policy (OSTP) memo stipulated that all federal agencies must have policies in place to increase access to digital research data (OSTP, 2013). The OSTP memo highlights the importance of providing open access to datasets and scholarly publications as a method of promoting innovation, accountability, efficiency, and impact.

Academic journal editors and publishers are also stakeholders that can encourage researchers to make data accessible. While not all researchers are funded by federal agencies, most wish to publish their findings in a reputable journal. Thus, the policies that academic journals have regarding data access may affect the likelihood that researchers will provide access to their data. There appears to be a growing trend among leading scientific journals to require that the data supporting publications be made publicly available as part of the publication process. A key example of this move within the journal community is the Joint Data Archiving Policy (JDAP), which has been adopted by a variety of scientific journals particularly within the fields of evolutionary biology (Dryad, 2011).

Dryad, a repository created to preserve scientific data, supports the JDAP initiative by providing the necessary infrastructure for data sharing. The repository has created a system for authors to submit their data during the manuscript submission process. This system also provides a unique and stable data citation and creates a persistent link between the article and the underlying data. While other authors have examined the effects of journal policies to increase data availability and sharing in scientific fields like evolutionary biology, there has not been similar work performed in the social sciences.

This paper explores the relationship between social science journal data sharing policies and research data availability. The purpose of this study is to gain a greater understanding of how social science journals can positively affect public availability to research data through their data sharing policies. The following research questions inform the study:

- What are the journal data sharing policies for the top 100 social science journals?
- Is the data underlying social science articles published in 2011 available online?
- Where is the data available online (i.e., data repository/center, institutional repository, personal website, as a supplementary document to the article)?
- Is there a relationship between social science journal data sharing policies and accessibility?

Determining the impact of data sharing policies on data accessibility may encourage other social science journals to implement or strengthen their data sharing policies. This in turn could help increase the amount of social science research data available for replication and secondary analysis and ultimately affect transparency, innovation, and the efficient use of public funds.

CHAPTER TWO

LITERATURE REVIEW

Over the past ten years there has been a data sharing movement advocating for the value of research data and creating the infrastructure for archiving and sharing data. An influential work by Van de Sompel, Payette, Erickson, Lagoze, and Warner (2004) called on the scholarly community to expand its definition of a “unit of scholarly communication” to include other materials, including datasets. The authors stated that this “future scholarly communication system” must also include mechanisms for the preservation and early registration of materials. Today there are infrastructures in place for preserving datasets, and there has been significant progress towards viewing data as a “unit of scholarly communication” in its own right. An example of this shift is a recent change to the NSF grant proposal guide. NSF now allows researchers to list datasets on their Biographical Sketch and renamed the “Publications” section to “Products” (NSF, 2013).

Despite the progress over the past ten years, the sharing of research data is still not a consistent practice (Borgman, 2012). To gain a more holistic picture of the current state of data sharing the following topics are examined below: (1) the benefits of sharing, (2) researchers’ perceptions, (3) the infrastructure for sharing and preservation, (4) the incentives to share, (5) the dimensions of sharing social science data, and (6) the effect of data sharing policies.

Benefits of Sharing

The benefits of sharing research data have been widely discussed within the literature. The sharing of research data is essential for the replication and verification of results, which is a key aspect of the scientific process (King, 2006). Sharing also supports transparency and accountability for research findings and by extension lessens potential fraud (Piwowar & Chapman, 2010). Sharing of data can also reduce redundant data collection promoting efficient use of money and time (Callaghan et al., 2012). In addition, if research is publicly funded then sharing increases the return on the public's investment and serves the public good (OSTP, 2013). Data availability also supports collaboration within and across disciplines, increasing the potential for different interpretations and uses of the data (Callaghan et al., 2012; Tenopir et al., 2011). Finally, datasets are important pedagogical tools for both undergraduate and graduate students and can provide them first-hand experience with replicating findings (King, 2006).

Researchers' Perceptions

Authors have largely used surveys to examine researchers' perceptions about sharing research data. A report sponsored the Research Information Network (RIN) examined why researchers do or do not share data. The report by Swan and Brown (2008) found that researchers cited a lack of experience with data management, a lack of expertise in making data accessible through the inclusion of comprehensive methodological information and metadata, and a lack of knowledge on where to archive the data. The report also found that researchers often fear that their data will be "hijacked" or they will be "scooped" by sharing their data, pointing to the highly competitive nature of academia (p. 28).

Another significant barrier involves the investment of time and money necessary to share one's data. A study by Tenopir et al. (2011) found that 53.6% of respondents cited insufficient time and 39.6% cited lack of funding as their reasons for not making data available electronically (p. 7). Swan and Brown (2008) also cited time and money as key reasons researchers do not share. Likewise, both studies found that ownership rights issues affected sharing. However, Swan and Brown (2008) also found that most researchers had tried to share data with others who directly requested it but there were often issues related to insufficient metadata, insufficient time to prepare the data for sharing, or an inability to locate the requested data. Tenopir et al. (2011) reached the conclusion that the barriers to data sharing are deeply rooted within culture, researcher practices, and the research process. In short, scholars have found that, even when researchers wish to share their data, there are significant challenges related to lack of knowledge, career-related concerns, and logistical barriers.

Infrastructure for Sharing and Preservation

Sharing of research data can come in a variety of forms. Methods can be more informal, such as direct peer-to-peer requests, posting the data on a personal website, or including the data as a supplementary document accompanying an article. Methods can also be more formal such as archiving the data within data repositories or institutional repositories. Data sharing starts with an understanding of the data lifecycle, which is intertwined with the research lifecycle (Tenopir et al., 2011).

The Digital Curation Centre (DCC) provides a high-level overview of the stages involved in the lifecycle of digital material. This model can be used to “ensure the maintenance of authenticity, reliability, integrity and useability of digital material” (Higgins, 2008, p. 135). The DCC Curation Lifecycle Model (see Figure 1) shows how the management of

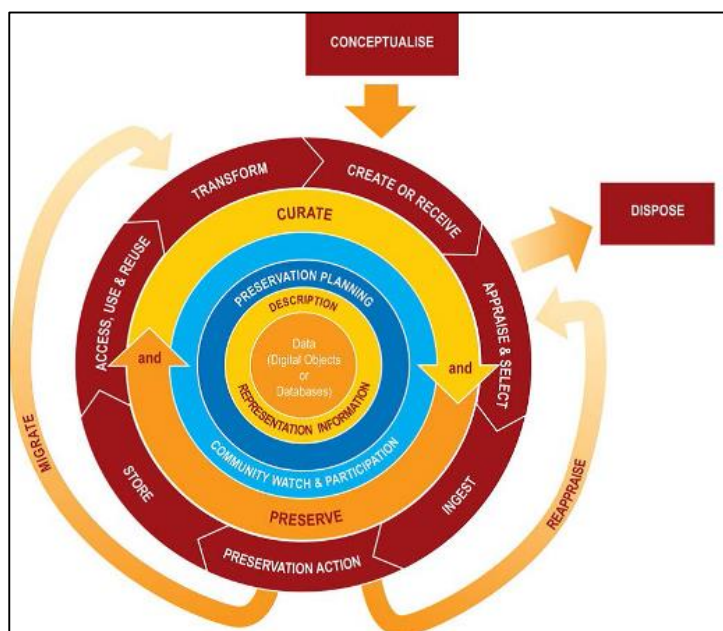


Figure 1. DCC Curation Lifecycle Model

data involves a conceptual understanding that planning is important at all lifecycle stages from the creation of the data to its reuse or disposal.

One aspect of ensuring data is reusable is the systematic capturing of methods during the research lifecycle as well the assignment of useful and informative metadata. Although a full discussion of the role of metadata in discovery and archiving is beyond the scope of this paper, it is important to note that metadata plays an important part in discovery and interoperability (Greenberg, White, Carrier, & Scherle, 2009).

Formal sharing involves storing data within a repository to ensure long-term preservation. Certain disciplines have addressed this need by suggesting that researchers deposit their data in discipline-specific repositories. For instance, Dryad was designed to preserve the data of evolutionary biology, ecology, and related disciplines (Greenberg, White, Carrier, & Scherle, 2009). Within the social sciences, the Data Preservation

Alliance for the Social Sciences (Data-PASS) was formed in 2004 to support the preservation of social science data and assisted with the creation of the Dataverse Network, an open-source application for publishing and accessing research data (Crosa, 2011). Other options for long-term preservation of research data may also include depositing the data in an institutional repository (Cragin, Palmer, Carlson, & Witt, 2010). Characteristics of these types of archiving institutions include explicit practices and written policies for ensuring long-term access to digital assets including the replication of copies off-site and procedures to address file format obsolescence (Pienta, Alter, & Lyle, 2010).

A final important piece for discovery and reuse is the presence of a data citation and ideally a link between the data and a published article. Mooney and Newton (2012) state that data citations should be a necessary part of data publication and reuse. However, based on a content analysis of journal articles, author instructions, style manuals, and data publishers, the authors found that “citation of data lags behind as a normative practice” (p. 13). This lack of standardization of data citations has been an area of advocacy and concentrated work among those calling for more extensive data sharing (Altman & King, 2007; Lawrence, Jones, Matthews, Pepler, & Callaghan, 2011). Data citations also relate fundamentally to the incentive structure for sharing research data.

Incentives to Share

Largely absent from the current academic culture are career-related incentives for sharing data. Researchers have stated that possible incentives could include standards for data citation, evidence of benefits of publishing datasets, and more explicit career rewards (Swan & Brown, 2008). In regard to data citations, DataCite has created

mechanisms for assigning Digital Object Identifiers (DOIs) to datasets. Likewise, the Dataverse Network creates persistent identifiers (Universal Numeric Fingerprints and handles) for datasets along with including other important elements like author, title, and URL (Altman & King, 2007). These and other initiatives are helping to create the necessary infrastructure for data citations. Nevertheless, maintaining data in a citable state and creating linkages to promote reuse involves communication and collaboration between data publishers, journal publishers, and authors (Ball & Duke, 2012). According to scholars, creating a data citation culture is important for incentivizing researchers to share by ensuring proper scholarly acknowledgement (Mooney & Newton, 2012).

The literature also addresses specific benefits of publishing datasets. Authors have found evidence of a significant association between publishing data and increased citations (Piwowar, Day, & Fridsma, 2007). Researchers have also mentioned that a reason for not sharing data is a lack of belief that others would wish to use their data (Swan & Brown, 2008). A study by Piwowar and Vision (2013) has found that a substantial percentage of archived gene expression microarray datasets are reused, and reuse continues for years after the original researchers finished publishing on the data. This finding is evidence of the benefits of data sharing for the scholarly community. Creating the structure for proper scholarly attribution for data and identifying specific career advantages are important aspects of increasing the rate of data sharing.

Sharing Social Science Data

Different disciplines have different norms and opinions concerning making data available to a broader public (Swan & Brown, 2008; Tenopir et al., 2011). Certain disciplines such as atmospheric science and environmental science and ecology appear

more willing to share their data than other disciplines. Disciplines such as medicine and the social sciences are more likely to restrict data access due to the use of human subjects (Tenopir et al., 2011). This variation suggests that factors related to measurement within one discipline may not necessarily apply to another.

Despite the ethical considerations that may accompany social science data there is a call within the literature to share data. The social sciences have seen a concentrated effort for some time to encourage data availability through software, policies, and researcher practices (King, 2011). King (2011) points out that within the social sciences there are now opportunities to “understand and address major previously intractable problems that affect human society” (p. 719). King goes on to discuss that certain mechanisms could be put in place to simultaneously protect privacy while also supporting data sharing (p. 719). Bishop (2009) makes an ethical argument for sharing qualitative data. She writes that researchers not only owe a duty to participants but also to the scholarly community and the public at large, which involves not duplicating research efforts and supporting an open and transparent scholarly community. In short, while social science data may present unique challenges, scholars believe there are important reasons to share this data.

The Effect of Data Sharing Policies

Many funders require some form of plan for sharing research data, but there are currently no formal penalties or enforcement mechanisms in place for non-compliance (Pienta et al., 2010; Tenopir et al., 2011). Professional organizations in the social sciences also have various ethical guidelines or “best practices” related to the sharing of research data including the American Sociological Association, American Psychological

Association, and the American Association for Public Opinion (Pienta et al., 2010). The field of economics has gone one-step further. Certain journals, including the American Economic Review, require that replication data must accompany any published article. Since enforcement is often not a part of these policies, investigation is important for measuring their effectiveness; as the saying goes “you cannot manage what you do not measure” (Piwowar, 2011).

Various studies have found a positive correlation between journals with strong data sharing policies and data availability. Piwowar (2011) used bibliometric methods to identify factors that are associated with researchers archiving raw gene expression microarray datasets after publication and found a positive association between data sharing and “a journal with a relatively strong data sharing policy” (Abstract section, para. 3). Likewise, Vines et al. (2013) examined articles within the field of evolutionary biology and found that journals with policies requiring a data statement significantly increased the rate of data availability online and raised the odds of retrieving the data twenty five times (p. 1305). Vines et al. (2013) also found a journal that recommends “data archiving is only marginally more effective than having no policy at all” (p. 1306).

Other studies have examined the willingness of researchers bound by a particular journal’s (PLoS) data sharing policy to provide data upon request and concluded that journal data sharing policies do not lead to researchers making their data available (Savage & Vickers, 2009). Alsheikh, Qureshi, Al-Mallah, Ioannidis, and Boutron (2011) also examined the level of compliance with data sharing policies within high impact journals. They found a relatively high level of non-compliance (59%), and only 9% of researchers deposited their data online.

There is less research analyzing policies, especially journal data sharing policies, within the social sciences. A paper presented by Pienta et al. (2010) at the Organisation, Economics and Policy of Scientific Research Conference constructed a massive database of social science studies over the last 40 years and specifically examined the availability of social science data from research that NIH or NSF funded. The authors found that most social science data is not formally shared through a data or institutional repository (11.5%), half the data funded by awards are only shared informally (44.6%), and almost half is only shared among the research team (43.9%).

These studies show a lack of consensus within the literature on the effectiveness of journal data sharing policies. They also show that a large amount of the research on journal data sharing policies has been based within a particular field (evolutionary biology), or focused on a particular journal (PLOS), or a particular type of journal (high impact journals). All the studies are consistent in showing that the rate of data sharing is relatively low and a high variability of compliance exists among researchers. The literature presents an opportunity for expanding the discussion surrounding journal data sharing policies to the social sciences. The growing push by external stakeholders to increase the rate of sharing also suggests that this field of study necessitates continuous measurement and research to help improve practice and expand the knowledge of the effectiveness of certain techniques.

CHAPTER THREE

METHODS

The research design of this study is quantitative and exploratory in nature. This study gathered documentary evidence in the form of journal data sharing policies, journal articles, and the artifacts of electronically available data. Informetrics, a variety of bibliometrics, was used to analyze the relationship between data policies and accessibility. Informetrics involves the “use and development of a variety of measures to study and analyse several properties of information in general and documents in particular” (Osareh, 1996, p. 148). These methods are appropriate for this study based on the use of bibliometrics in similar studies (see Piwovar, 2011) and due to the researcher’s belief that the data gained through unobtrusive measures will provide a more accurate picture than self-report data from researchers (Wildemuth, 2009, p. 158).

The population for this study is social science journals. A preliminary sample was taken from the Thomson Reuters’ 2011 Journal Citation Reports (JCR) that includes the top 100 most cited social science journals based on impact factor. Thomson Reuters’ JCR uses a broad definition of social sciences; some of the journals in the sample also fall in the category of health sciences (see Appendix C for a full list of the top 100 social science journals). The Journal Research Data Policy Bank (JoRD) project sponsored by Jisc (Joint Information Systems Committee) has assembled a list of social science

journals with some form of research data policy (JoRD, n.d.). The JoRD list was used to determine which of the top 100 social science journals previously identified have a research data sharing policy.

The journals with policies were then examined and three data sharing policy categories were created: (1) mandatory, (2) recommended, (3) no data sharing policy, although supplementary materials may be supported. Any journal that JoRD had not found to have a policy was automatically added to the no policy category. Then six journals were randomly selected from the recommended category and from the no policy category. Only one journal had a mandatory data sharing policy, so this journal was automatically added to the sample.

The second step of the data collection procedures was the identification of five articles from each journal for analysis. Articles were randomly selected and then checked to determine whether data was used in the analysis. If data was not used (e.g., systematic reviews, editorial pieces, book reviews, commentaries, etc.) then that article was removed from the sample, and another article was randomly selected. This process continued until five articles from each journal had been selected, totaling 65 articles. Randomization protects the article authors from potential deductive disclosure and assists in establishing internal validity for the study.

The final step of the data collection procedures involved collecting the information on data availability for the 65 articles. This involved an in-depth online structured search for the data (see Appendix A for information on the search strategy). After the availability of the data was determined, each article was coded with a one (data is available) or a zero (data is not available). Data availability was defined in the broadest

of terms (i.e., the data underlying the article could be found in some form) but the quality of the data was not evaluated. Data was also not required to be available in the final form used in the analysis for it to be coded as available. For instance, if an author used two different datasets and merged these datasets for analysis, the data was coded as available if both datasets could be found but not the merged dataset. Information was also collected and coded on (1) whether the data was existing or original data, (2) the data type (i.e., quantitative, qualitative, survey, mixed, or other), and (3) the location of the data.

CHAPTER FOUR

RESULTS

Policy Types

The 100 journal policies were placed within one of three broad categories (i.e., mandatory, recommended, or no policy). Some variability was present among these policy types, especially in the recommended category. For instance, some journals required authors to make data available upon request, while others strongly suggested that authors make data publicly available. Other journals used a professional organization's ethical guidelines to encourage researchers to share their data. Likewise, in the no policy category some journals state authors can include data as supplementary materials, which allows researchers to store and make accessible data through the journal's website. However, no normative statements or recommendations to share data were made by these journals outside of making the infrastructure available to share. Table 1 provides example language for these three policy types (see Appendix B for the data sharing policies for the 13 journals within the sample).

Table 1. Journal Data Sharing Policy Examples

Policy	Example Text
Mandatory Data Replication	It is the policy of the <i>American Economic Journal</i> to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the <i>American Economic Journal</i> , prior to publication, the data, programs, and other details of the computations sufficient to permit replication. (American Economic Journals, n.d.)
Recommended Data Sharing	<p><i>Making data publicly available strongly encouraged</i></p> <ul style="list-style-type: none"> • If the study includes original data, at least one author must confirm that he or she had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis. We strongly encourage authors to make their source data publicly available. (British Journal of Psychiatry, 2014, Access to data section) <p><i>Sharing encouraged based on ethical guidelines</i></p> <ul style="list-style-type: none"> • All persons who publish in American Sociological Association (ASA) journals are required to abide by ASA guidelines and ethics codes ... This requirement includes adhering to ASA's stated policy on data-sharing: "Sociologists make their data available after completion of the project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the data or the anonymity of research participants (e.g., raw field notes or detailed information from ethnographic interviews)" (ASA Code of Ethics, 1997). (American Sociological Review, n.d., Ethics section)
No Policy (although supplementary materials may be supported)	Elsevier accepts electronic supplementary material to support and enhance your scientific research. Supplementary files offer the author additional possibilities to publish supporting applications, high-resolution images, background datasets, sound clips and more (Cognitive Psychology, 2014, Supplementary data section).

Only one journal within the top 100 social science journals was found to have a mandatory policy, the *American Economic Journal: Macroeconomics*. Twenty journals fell into the recommended category and 14 of these journals based their recommendation on a professional organization's ethical guidelines. The table below details the distribution of data sharing policies among the 100 journals.

Table 2. Distribution of Data Sharing Policies

Policy	Count	Percent
Mandatory	1/100	1%
Recommended	20/100	20%
No Policy	79/100	79%

Data Types

Of the 65 journal articles examined, nine articles performed secondary analysis on existing data, whereas 56 articles used original

Table 3. Existing vs. Original Data

Data	Count	Percent
Existing	9/65	14%
Original	56/65	86%

data in some form for their analysis. Information on whether the authors used qualitative data, quantitative data (i.e., randomized control trials, laboratory experiments, etc.), survey data, mixed data, or other forms of data (i.e., simulations and economic models) was also collected. Table 4 displays the types of data found in the sample.

Since surveys can contain both quantitative and qualitative elements based on the methodology, surveys were coded separately. The majority of the articles used quantitative data (approximately 61%). Surprisingly, only one article solely used qualitative data. However, out of the six articles that used mixed data types, three did use qualitative data in the form of interviews or observations and four used surveys in conjunction with other data types.

Table 4. Data Types of 65 Articles

Data Type	Count	Percent
Quantitative	40/65	61%
Qualitative	1/65	2%
Survey	9/65	14%
Mixed	6/65	9%
Other (simulations, models, etc.)	9/65	14%

Data Availability: Existing vs. Original Data

Eight out of nine (89%) of the existing datasets were located online. However, for two of these articles the authors used merged datasets in their analysis where the authors combined multiple existing datasets to answer a research question. Although the individual publicly available

Table 5. Data Availability

Data	Count	Percent
Existing	8/9	89%
Original	8/56	14%
Total	16/65	25%

datasets were located, the merged datasets were not available. Three of these articles used large-scale national surveys such as the National Morbidity, Mortality and Air Pollution Study. Of the 56 articles using original data, data was located for eight articles (14%). For the entire sample, 25% of the underlying data was discovered online.

Data Availability: Journal Policies, Location, and Data Types

For the journal articles with a mandatory data replication policy, 100% of the data was available online. Of the eight articles for which original data was located, the majority (63%) came from the five articles with a mandatory data replication policy (see Table 6).

Table 6. Original Data Availability by Policy and Location

Journal Data Policy	Count of Available Original Datasets	Percent	Location
Mandatory	5/8	62.5%	Supplementary material
Recommended	2/8	25%	Supplementary material, Other
No Policy	1/8	12.5%	Data archive

The data from the articles required to comply with the mandatory data replication policy was located alongside the article under “Additional Materials” with a hyperlink to “Download Data Set” within the online journal portal (see Figure 2).

The screenshot shows a journal website interface with the following sections:

- Article Citation:** A blue header bar followed by a greyed-out text area.
- DOI:** A label followed by a greyed-out text area.
- Abstract:** A blue header bar followed by a large greyed-out text area.
- Article Full-Text Access:** A blue header bar followed by a red link labeled "Full-text Article".
- Additional Materials:** A blue header bar followed by two red links: "Download Data Set (64.42 KB)" (which is circled in red) and "Online Appendix (121.76 KB)".

Figure 2. Data Included as Additional Materials on Journal Website

For the two datasets within the recommended category, one dataset came from a study performed by a large-scale multi-study research group. The group’s website stated that the data and supporting documentation will be published at the Inter-university

Consortium for Political and Social Research (ICPSR) data archive. The research group also provided contact information for gaining access to the data prior to its publication on ICPSR. Although this data was technically not “available” online, the message about the data becoming available through a data archive in the future and the contact information for procuring the data was determined by the researcher to be sufficient proof that the data is available for all intents and purposes.

The data in the no policy category also came from a large-scale research project. The researcher is an economist who used the Dataverse Network (DVN) to share his data. The DVN has the archival infrastructure to preserve data and create a data citation that includes a unique identifier. The DVN also allows individual scholars to create their own Dataverses to store and share their research data. This researcher’s Dataverse included various data types from GIS data to survey data with a total of 40 studies in the Dataverse.

The other available data also varied in data types (see Table 7). Notably, the other article that used mixed data shared their data as supplementary materials on the journal website. All the various data types were included in a Zip file with a README.rtf file that explained that: “The data comes from many different sources, all listed in the paper. Hence, there are many different files.” The three original datasets that fell into the other category were economic models or simulations that were shared as supplementary materials. In comparison, the existing data was less varied in type than the original data and primarily included surveys and quantitative data in the form of large-scale publicly available datasets, such as birth and death files or income inequality data.

Table 7. Data Types of Available Data

Data Type	Existing Data		Original Data		Total	
	Count	Percent	Count	Percent	Count	Percent
Quantitative	5/8	62.5%	2/8	25%	7/16	44%
Survey	3/8	37.5%	1/8	12.5%	4/16	25%
Mixed	0	0%	2/8	25%	2/16	12%
Other	0	0%	3/8	37.5%	3/16	19%

Other Findings

Although the willingness of researchers to share their data through other methods, such as peer-to-peer sharing, was not examined in this study, it is of note that in one article the authors explicitly stated in the notes section that “data and material are available upon request.” This was found in a journal that had a recommended data sharing policy based on ethical guidelines. In further studies it would be worthwhile to determine whether authors that consciously accompany articles with such statements are more willing to share their data informally than others.

CHAPTER FIVE

DISCUSSION & CONCLUSION

Policy Types

This study found that only one of the 100 journals had a mandatory policy where researchers had to provide replication data prior to publication. While some journals did make statements related to sharing data upon request, making data publicly available, and the ethical imperative to share, most journals did not make strong normative statements about the importance or need to share data. These findings suggest that using journal data sharing policies as a mechanism to influence or increase the rate of data sharing is currently not a prominent practice within social science journals. This also points to an opportunity for journals to revise their data sharing policies if they wish to positively influence the rate of data sharing.

Currently there is a growing trend among some journals to strengthen existing data archiving and sharing policies. For instance, PLOS put into effect a revised data policy on March 1, 2014 that requires authors to include a data availability statement with all published articles (PLOS, 2013). This trend suggests that some journals are viewing mandated data sharing policies as an important route to increase the rate of data sharing.

Data Types

This study found that the majority of articles (86%) used original data with only 14% of authors using existing data. This is of note because of the articles that used existing data the researcher was able to locate 89% of those datasets. One of the rationales for increasing data sharing is the importance of replication and verification of results to support the scientific endeavor (King, 2006). If researchers use publicly available datasets (such as a large scale national survey) for their analysis then the ease of replicating and verifying these results increases. In addition, by using existing datasets, authors are increasing the transparency of their findings. This is not to say that it is preferable to use existing data versus original data. The purpose is to highlight the value created by data that is publicly available versus original data that has not been shared.

Different data types, from quantitative randomized control trials to interviews, surveys, GIS data, and observational data, also present different challenges and opportunities. It has been discussed in the literature that qualitative and health data raise specific confidentiality and ethical challenges because of the need to protect personally identifiable information (PII), and this in turn can impact the willingness and ability of researchers to share their data (Tenopir et al., 2011; Bishop, 2009). Of the available data found, none of the datasets included interview, observational, or focus group data. The affordances offered by different types of data raises another issue related to data sharing. Since all data is not created equal, some data may be easier to share than others.

For instance, the articles that used original data and fell into the other data type category primarily used models or simulations. The “data” that is created from these types of articles often comes in the form of replication code for the model or simulation. This type of data may involve less effort to make the data ready to share versus other data

types such as surveys, interviews, or health studies that may require de-identification, which can be a time-intensive process. The need to prepare data for sharing and the effort this takes speaks to the findings by Tenopir et al. (2011) that researchers often cite insufficient time or a lack of funding as a reason not to share. Raising the question whether funding bodies that require data sharing, such as NIH, should also require researchers to include resources for data management and archiving in budget documents.

Where to store and how to share mixed data types may also raise particular issues for researchers. In a recent Data Pub blog post, Strasser (2014) discusses that one potential limitation of discipline-specific data repositories is that they may only accept certain data types. Therefore, researchers may be unable to store mixed data types from one study in a single data repository. Strasser writes that this can be a strength of using an institutional repository (IR) because it can often store and make accessible the various types of data from an academic project. Strasser also suggests that researchers may want to consider using both a data repository and institutional repository. As she writes, “selecting a repository for your data doesn’t need to be either an IR or discipline-specific repository (DR). These repositories each have advantages and disadvantages, so using both makes sense” (IRs versus discipline-specific repositories section).

In this study, it was found that a researcher self-archived mixed data in the DVN. The DVN does not have restrictions on the types of data one can archive, but it does have a two gigabyte file size limitation for each file uploaded, which could cause issues with sharing large datasets (DVN, n.d.). No researchers were found to have used an IR to share their data, although another potential benefit of institutional repositories is that an article and the underlying data could be made discoverable at a single location, which

“helps ensure reproducibility and transparency” (Strasser, 2014, IRs: the whole enchilada section). Essentially, data types influence the effort involved in sharing data as well as where the data can be effectively stored, preserved, and accessed.

Data Availability

The findings from this study suggest that the only type of journal data sharing policy that impacts the rate of sharing is mandatory policies. This finding is in-line with previous literature that found a correlation between strong data sharing policies and data availability (Piwowar, 2011; Vines et al., 2013). Likewise, of the three articles without a mandatory policy, two came from journals with recommended policies and one came from a journal with no policy, echoing Vines et al.’s (2013) findings that a “recommended” policy only makes a marginal difference over no data sharing policy.

This suggests that if journals wish to positively impact the rate of data sharing they should implement mandatory data sharing policies. These policies could require authors to prove their data is publicly available in the form of a data accessibility statement prior to publishing the article. If the data sharing community wants to significantly increase the rate of sharing then enforcement mechanisms should also be further examined and improved.

This study also found a low percentage of original data (14%) being shared online. Only 5% of the total articles published in journals without a mandatory policy shared their data, which is in-line with Pienta et al.’s (2010) findings related to the low number of social science researchers who share their data. It is also of note that the only journal with a mandatory policy was an economics journal, and one of the three researchers who shared their data without a mandate was also an economist. This

highlights how data sharing norms are often discipline-specific. As Pienta et al. (2010) points out, even though certain social sciences disciplines, such as economics and political science, began data sharing efforts early in the data sharing movement, there is still a large amount of “heterogeneity in data sharing in the social sciences” (p. 2). This suggests that social science disciplines interested in increasing the rate of data sharing can use other disciplines, such as economics, as a model and increase inter-disciplinary data sharing discussions.

Two of the three datasets that did not fall under the mandatory data replication policy also came from large-scale research projects. This raises the question: does the scale of a project affect researchers’ ability or willingness to make data available? This potential relationship also ties back to resources, since larger research projects may have more funds available to dedicate time and money to preparing data to share. The association between the size of research projects and data sharing could be examined in future work.

Data Location

The infrastructures and methods researchers use to share and store their original research data has various implications. Since the *American Economic Journal: Macroeconomics* requires authors to provide their replication data to the editor prior to publication, it is not surprising that the data is made available as supplementary materials to the article. The data being placed alongside the article significantly eases access and makes it explicitly clear that this is the underlying data.

However, this system of storing the data on a publisher’s or journal’s servers also raises certain preservation concerns. As discussed in the literature review, there are a

variety of ways that researchers can make their data available; however, more formal methods for sharing, such as data repositories, can support the long-term preservation of digital data assets. Storing one's data on a commercial publisher's site may not involve archival preservation activities or the assignment of metadata, which aids in discovery.

Likewise, the data stored as supplementary materials on the *American Economic Journal: Macroeconomics* website includes no data citation. Data citations are an important aspect of establishing data as a standalone "unit of scholarly communication" (Van de Sompel et al., 2004). They also provide scholarly acknowledgement to authors, which can be used as an incentive to encourage researchers to share their data (Mooney & Newton, 2012). Therefore, although these datasets are easily available as supplementary materials they are not ensured long-term preservation and do not contain a data citation that encourages acknowledgement.

Two of the original datasets were, or were soon to be, available through a data archive. A data archive, like the DVN or ICPSR, provides a strong data preservation and data citation infrastructure for a study. However, discovery and accessibility can be an issue, especially if a user is specifically looking for the data underlying a scholarly article. During this study significant time and effort was required to locate the dataset available in the DVN. For this reason, there has been a push within the data sharing movement to create persistent links between a published article and the underlying published data (Wynholds, 2011). Establishing a link between an article and the data stored in a repository can assist with the following:

- Facilitate other researchers to find data without significant outside investigation
- Promote reuse of data by increasing discoverability

- Promote the awareness of the value of data as a significant stand-alone research product
- Increase the general public's awareness of the value of research data through increasing the visibility of data citations

However, to make this link will require cooperation between data repositories, journals, and publishers (Ball & Duke, 2012).

Multiple grant-funded projects are already working to integrate data publication within the journal manuscript submission system with an end result of creating a link between the underlying data and a scholarly article. For instance, Dryad has created a workflow to facilitate the archiving and linking of data and makes this service available to journals at no cost (Dryad Submission Integration, 2013). Likewise, the Public Knowledge Project-Dataverse Integration Project is a collaboration aimed at integrating manuscript and data submissions through the Open Journal Systems (OJS) and the Harvard Dataverse Network.

These two projects are examples of work toward linking data and articles. However, continuing this work requires investment in the data repositories partnering with journals, performing the archival tasks, providing the infrastructure, facilitating data discovery through metadata, and advocating for data sharing. For instance, Vines et al. (2013) cites a recent study that found that the cost of running the Dryad database is around \$400,000 a year. Despite the increasing need for data repositories that preserve and provide long-term access to data, repositories continue to face financial challenges.

A recent call for a change in the funding streams for domain repositories discusses these challenges. This statement describes how there is currently a growing

need for data repositories because of federal mandates, such as the OSTP memo, and the increasing momentum of the data sharing movement (ICPSR, 2013). However, these same repositories often rely on project-based grants, which do not guarantee long-term sustainable funding. The call concludes that while there may not be a single solution, “*creating sustainable funding streams will require the coordinated response of multiple stakeholders in the scientific, archival, academic, funding, and policy communities*” (A call for change section). In short, increasing access to digital research data will require cooperation between stakeholders and an understanding that effectively sharing research data involves significant resources.

Limitations

This study did not collect information on a variety of variables that could affect a researcher’s decision to share their data including the funding agency, the length or scope of the project, and the experience of the researcher. As previously discussed, there are many factors that can affect a researcher’s choice to share or not to share; however, collecting information on these other factors was outside the scope of this study. Since this study is exploratory, the findings are not generalizable.

It was also assumed that the journals with higher impact factors would be leading others in data sharing as found by Piwowar and Chapman (2010); however, this means the findings do not give a complete picture of data sharing across all social science journals. This study also did not examine informal peer-to-peer data sharing, which may result in an underestimation of data sharing. It is important to note that although the search strategy was designed to be extensive, it is possible data existed online that was not discovered, potentially underestimating data sharing.

Future Work

As discussed above some policies stated that authors are required to share data upon request and in one article authors explicitly stated that data was “available upon request.” A next step in this study would be to attempt to contact researchers directly and request the data. This would also allow a comparison to other findings that have examined the impact of journal data sharing policies on peer-to-peer requests (see Savage & Vickers, 2009; Vines et al., 2013). Likewise, expanding this study would create useful benchmarks for understanding social science researchers’ willingness and ability to share data informally.

This dataset could also be used to examine whether there is a correlation between the impact factor of a given journal and its data sharing policy type and data availability. Finally, although a random sample was taken to explore broadly social science journals data sharing policies in conjunction with data availability, it would be interesting to use purposive sampling of the journals and chose journals within specific social science disciplines. For instance, one could select a certain number of sociology, economics, political science, and psychology journals and then replicate the study to examine how various social science disciplines compare.

Conclusion

Individuals within the data sharing movement have made various arguments for why sharing research data is important. These arguments include the importance of replicating and verifying results, supporting the transparency and accountability of research findings, increasing the public return on investment by allowing secondary analysis and reuse, and using data as a pedagogical tool. However, the rate of sharing is

still not consistent or substantial. Methods to increase the number of researchers who share their data can take a variety of forms. Some scholars are pointing to career advantages, such as increased citation rates. Other stakeholders, such as journal editors, are implementing data sharing policies to encourage researchers to share their data.

This study examined journal data sharing policies within the social sciences to more fully understand the relationship between these policies and data availability and found that only mandatory data sharing policies impact data sharing rates. In addition only a small number of researchers chose to share their research data. Journals can use these findings to inform decisions on maintaining, strengthening, or more strictly enforcing policies.

Where researchers choose to share their data also has various implications. Data repositories support the long-term preservation of digital data as well as encourage acknowledgement of data through the use of data citations. However, to make research data easily accessible a link should be established between the data stored in a repository and the journal website. Creating these data citation links would also further incentivize researchers to share through increasing the visibility of their data as a standalone research product. This requires cooperation between various stakeholders and an investment in the long-term future of data repositories to ensure a robust data access and preservation system is built for future generations.

REFERENCES

- Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLoS ONE*, *6*(9), e24357. doi:10.1371/journal.pone.0024357
- Altman, M., & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine*, *13*(3/4).
- American Economic Journals: Data Availability Policy. (n.d.). Retrieved from <http://www.aeaweb.org/aej/data.php>
- American Sociological Review. (n.d.). Manuscript submission. Retrieved from <http://www.uk.sagepub.com/journalsProdDesc.nav?prodId=Journal201969#tabview=manuscriptSubmission>
- Ball, A., & Duke, M. (2012). *Data citation and linking* (Briefing Paper). Edinburgh, UK: Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking>
- Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, *44*(3), 255–272.
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. doi:10.1002/asi.22634
- British Journal of Psychiatry. (2014). Instructions for authors. Retrieved from <http://bjp.rcpsych.org/site/misc/ifora.xhtml>

- Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., Wright, D. (2012). Making data a first class scientific output: Data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation*, 7(1). doi:10.2218/ijdc.v7i1.21
- Cognitive Psychology. (2014). Guide for authors. Retrieved from <http://www.elsevier.com/journals/cognitive-psychology/0010-0285/guide-for-authors#87000>
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023–4038. doi:10.1098/rsta.2010.0165
- Crosas, M. (2011). The Dataverse Network®: An open-source application for sharing, discovering and preserving data. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-crosas
- DataCite. (n.d.). What do we do? Retrieved from <http://www.datacite.org/whatdowedo>
- Data-PASS. (n.d.). About. Retrieved from <http://www.data-pass.org/about.jsp>
- Dataverse Network. (n.d.). Creating new studies. Retrieved from <http://guides.thedata.org/book/create-new-study>
- Dryad. (2011). Joint data archiving policy. Retrieved from <http://datadryad.org/pages/jdap>
- Dryad. (2013). Submission integration overview. Retrieved February 26, 2014 from the Dryad Wiki: http://wiki.datadryad.org/Submission_Integration:_Overview

- Emotion. (n.d.). Instructions for authors. Retrieved from
<http://www.apa.org/pubs/journals/emo/index.aspx>
- Epidemiology. (n.d.). Instructions for authors. Retrieved from
<http://edmgr.ovid.com/epid/accounts/ifauth.htm>
- Greenberg, J., White, H. C., Carrier, S., & Scherle, R. (2009). A metadata best practice for a scientific data repository. *Journal of Library Metadata*, 9(3-4), 194–212.
doi:10.1080/19386380903405090
- Hey, T., & Trefethen, A. (2003). The data deluge: An e-science perspective. In F. Berman, G. Fox, & T. Hey (Eds.), *Wiley series in communications networking & distributed systems*. Chichester, UK: John Wiley & Sons, Ltd.
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. doi:10.2218/ijdc.v3i1.48
- Inter-university Consortium for Political and Social Research (ICPSR). (2013, June 24-25). Sustaining domain repositories for digital data: A call for change from an interdisciplinary working group of domain repositories. Retrieved from
<http://datacommunity.icpsr.umich.edu/sustaining-domain-repositories-digital-data-call-change-interdisciplinary-working-group-domain>
- JAMA Psychiatry. (2014). Instructions for authors. Retrieved from
<http://archpsyc.jamanetwork.com/public/instructionsForAuthors.aspx#SecEditorialPoliciesforAuthors>
- Jisc. (2012). Journal research data policy bank. Retrieved from
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/research-data-publication/jord.aspx.

- JoRD. (n.d.). Social science journals that have a research data policy. *JoRD blog*.
Retrieved from <http://jordproject.wordpress.com/project-data/social-science-journals-that-have-a-research-data-policy/>
- Journal of Marketing. (n.d.). AMA journals – editorial policies and procedures. Retrieved from <https://www.ama.org/publications/Pages/ama-editorial-policy-journals.aspx>
- King, G. (2006). Publication, publication. *Political Science and Politics*, 39(1), 119–125.
- King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331(6018), 719–721. doi:10.1126/science.1197872
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and peer review of data: moving towards formal data publication. *International Journal of Digital Curation*, 6(2). doi:10.2218/ijdc.v6i2.205
- Mooney, H., & Newton, M. (2012). The anatomy of a data citation: Discovery, reuse, and credit. *Journal of Librarianship and Scholarly Communication*, 1(1).
doi:10.7710/2162-3309.1035
- National Institutes of Health (NIH). (2003). *Final NIH statement on sharing research data* (No. NOT-OD-03-032). Bethesda, MD: National Institutes of Health.
Retrieved from <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html>
- National Science Foundation. (n.d.). *Dissemination and sharing of research results*.
Retrieved from <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- National Science Foundation. (2013). Grant proposal guide summary of changes.
Retrieved from
http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp

- Office of Science and Technology Policy. (2013). *Increasing access to the results of federally funded scientific research memorandum*. Washington, D.C.: Executive Office of the President. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Office of Management and Budget. (1999). *Uniform administrative requirements for grants and agreements with institutions of higher education, hospitals, and other non-profit organizations* (OMB Circular A-110, 36(d)(2)(i)). Retrieved from http://www.whitehouse.gov/omb/circulars_a110#36
- Osareh, F. (1996). Bibliometrics, citation analysis and co-citation analysis: A review of literature I. *Libri*, 46, 149-158.
- Pienta, A. M., Alter, G. C., & Lyle, J. (2010). *The enduring value of social science research: The use and reuse of primary research data*. Paper presented at the The Organisation, Economics and Policy of Scientific Research, Torino, Italy.
- PLOS. (2013, December 12). Data access for the open access literature: PLOS's data policy. Retrieved from <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
- Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*, 6(7), e18657.
doi:10.1371/journal.pone.0018657
- Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics*, 4(2), 148–156.
doi:10.1016/j.joi.2009.11.010

- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3), e308.
doi:10.1371/journal.pone.0000308
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1. doi:10.7287/peerj.preprints.1
- Public Knowledge Project-Dataverse Integration Project. (n.d.). About the project.
Retrieved from <http://projects.iq.harvard.edu/ojs-dvn/about-project>
- Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE*, 4(9), e7078.
doi:10.1371/journal.pone.0007078
- Strasser, Carly. (2014, February 20). Institutional repositories: Part 2. Retrieved from <http://datapub.cdlib.org/2014/02/20/institutional-repositories-part-2/>
- Swan, A., & Brown, S. (2008). *To share or not to share: Publication and quality assurance of research data outputs*. Research Information Network.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., Frame, M., & Neylon, C. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE*, 6(6), e21101. doi:10.1371/journal.pone.0021101
- Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking scholarly communication. *D-Lib Magazine*, 10(9).
doi:10.1045/september2004-vandesompel
- Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, 27(4), 1304–1308. doi:10.1096/fj.12-218164

Wildemuth, B. M. (2009). *Application of social research methods to questions in information and library science*. Westport, CT: Libraries Unlimited Press.

Wynholds, L. (2011). Linking to scientific data: Identity problems of unruly and poorly bounded digital objects. *International Journal of Digital Curation*, 6(1).

doi:10.2218/ijdc.v6i1.183

APPENDIX A: DATA AVAILABILITY SEARCH STRATEGY

1. Check the supplementary documentation accompanying the article on the journal website.
2. Search discipline-specific and/or location-specific data repository catalogs (i.e., ICPSR, the Dataverse Network, UK Data Archive, etc.) using article citation information.
3. If the article is connected to a larger research group or named study, search for the research group or specific study. Then examine the web presence of the research group or specific study looking for datasets.
4. Perform a general web search using variations of article title information and “data” or “datasets.” Discontinue reviewing results after 4 pages.
5. Perform a general web search for each article author, if authors have a personal website examine website for datasets or links to other data storage locations.
6. Determine whether the authors work at an institution with an institutional repository, if so, search the repository for the data.

APPENDIX B: JOURNALS AND DATA SHARING POLICIES

Journal	Policy	URL
Mandatory Data Replication Policy		
<i>American Economic Journal: Macroeconomics</i>	It is the policy of the American Economic Journal to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the AER Web site. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.	http://www.aeaweb.org/aej/data.php
Recommended Policy		
<i>The British Journal of Psychiatry</i>	If the study includes original data, at least one author must confirm that he or she had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis. We strongly encourage authors to make their source data publicly available.	http://bjp.rcpsych.org/site/misc/ifora.xhtml
<i>JAMA Psychiatry (previously Archives of General Psychiatry)</i>	If requested, authors should be prepared to provide the data and must cooperate fully in obtaining and providing the data on which the manuscript is based for examination by the editors or their assignees.	http://archpsyc.jamanetwork.com/public/instructionsForAuthors.aspx#SecEditorialPoliciesforAuthors

<i>Epidemiology</i>	<p>The editors encourage authors to provide information that enables other researchers to replicate their analyses (see editorial). For example, include:</p> <ul style="list-style-type: none"> • Counts for cell numbers in crude analyses • Analytic code used for the analysis of publicly available data • Code used to develop and analyze simulation data • Source of data, if publicly available 	http://edmgr.ovid.com/epid/accounts/ifauth.htm
<i>American Sociological Review</i>	<p>All persons who publish in ASA journals are required to abide by ASA guidelines and ethics codes regarding plagiarism and other ethical issues. This requirement includes adhering to ASA's stated policy on data-sharing: "Sociologists make their data available after completion of the project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the data or the anonymity of research participants (e.g., raw field notes or detailed information from ethnographic interviews)" (ASA Code of Ethics, 1997).</p>	http://www.uk.sagepub.com/journalsProdDesc.nav?prodId=Journal201969#tabview=manuscriptSubmission
<i>Journal of Marketing</i>	<p>The American Marketing Association (AMA) is committed to fostering the meaningful exchange of information to help create an environment for constructive criticism and free exchange of ideas. As publisher of the <i>Journal of Marketing</i>, <i>Journal of Marketing Research</i>, <i>Journal of International Marketing</i>, and <i>Journal of Public Policy & Marketing</i>, the AMA expects authors to adhere to the highest standards of integrity in research and the communication of research results and findings. Papers submitted to AMA journals should include enough information (including in-text, Web appendix, or other online supplements) so as to allow a reasonably trained researcher to replicate the results. This should include a precise description of the research and analysis procedures.</p>	https://www.ama.org/publications/Pages/ama-editorial-policy-journals.aspx

	The AMA requires the authors of manuscripts submitted to AMA journals to share additional details of their research findings and insights when requested by a journal editor. Although sharing and posting of data is not required, such additional information may include computer code, instruments, and other relevant information deemed necessary to facilitate replication. If there are any proprietary restriction on information, authors must notify the editor at time of manuscript submission.	
<i>Emotion</i>	APA Ethical Principles specify that “after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release” (Standard 8.14).	http://www.apa.org/pubs/journals/emotion/index.aspx
No Policy		
<i>Cognitive Psychology</i>	Elsevier accepts electronic supplementary material to support and enhance your scientific research. Supplementary files offer the author additional possibilities to publish supporting applications, high-resolution images, background datasets, sound clips and more.	http://www.elsevier.com/journals/cognitivepsychology/010-0285/guide-for-authors#87000
<i>Quarterly Journal of Economics</i>	Supporting material that is not essential for inclusion in the full text of the manuscript but would nevertheless benefit the reader can be made available by the publisher as online-only content linked to the online manuscript. The material should not be essential to understanding the conclusions of the article, but should contain data that is additional or complementary and directly relevant to the article content. Such information might include more detailed methods, extended data sets/data analysis, or additional figures.	http://www.oxfordjournals.org/our_journals/qje/for_authors/manuscript_instructions.html#Supplementary%20data

<i>Global Environmental Change</i>	Elsevier accepts electronic supplementary material to support and enhance your scientific research. Supplementary files offer the author additional possibilities to publish supporting applications, high-resolution images, background datasets, sound clips and more.	http://www.elsevier.com/wps/find/journaldescription.cws_home/30425/authorinstructions#N10C7E
<i>Journal of Finance</i>	N/A	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6261/homepage/ForAuthors.html
<i>Developmental Science</i>	We can also publish supporting information. Supporting information must be important, ancillary information that is relevant to the parent article but which does not or cannot appear in the main article. Supporting Information can comprise additional tables, data sets, figures, movie files, audio clips, 3D structures, and other related nonessential multimedia files. Like the manuscript accompanying it, it should be original and not previously published. If previously published it must be submitted with the necessary permissions.	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-7687/homepage/ForAuthors.html
<i>Journal of the American Geriatrics Society</i>	N/A	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1532-5415/homepage/ForAuthors.html

APPENDIX C: TOP 100 SOCIAL SCIENCE JOURNALS

The following list includes the top 100 social science journals as rated by impact factor according to the Thomson and Reuters' 2011 Journal Citation Report.

Journal Title	ISSN	Impact Factor
<i>Academy of Management Annals</i>	1941-6520	4.48
<i>Academy of Management Learning and Education</i>	1537-260X	4.8
<i>Academy of Management Journal</i>	0001-4273	5.608
<i>Academy of Management Perspectives</i>	1558-9080	3.75
<i>Academy of Management Review</i>	0363-7425	6.169
<i>Acta Psychiatrica Scandinavica</i>	0001-690X	4.22
<i>Addiction</i>	0965-2140	4.313
<i>Administrative Science Quarterly</i>	0001-8392	4.212
<i>Advances in Experimental Social Psychology</i>	0065-2601	4.889
<i>American Economic Journal: Macroeconomics</i>	1945-7707	3.8
<i>American Journal of Bioethics</i>	1526-5161	4.083
<i>American Journal of Geriatric Psychiatry</i>	1064-7481	3.638
<i>American Journal of Psychiatry</i>	0002-953X	12.539
<i>American Journal of Public Health</i>	0090-0036	3.926
<i>American Psychologist</i>	0003-066X	6.869
<i>American Sociological Review</i>	0003-1224	4.422
<i>Annals of Behavioral Medicine</i>	0883-6612	4.2
<i>Annual Review of Clinical Psychology</i>	1548-5943	9.111
<i>Annual Review of Environment and Resources</i>	1543-5938	6.419
<i>Annual Review of Psychology</i>	0066-4308	16.833
<i>Annual Review of Public Health</i>	0163-7525	5.451
<i>Annual Review of Sociology</i>	0360-0572	4.442
<i>Archives of General Psychiatry</i>	0003-990X	12.016
<i>Autism Research</i>	1939-3792	3.686
<i>Behavioral and Brain Sciences</i>	0140-525X	25.056
<i>British Journal of Psychiatry</i>	0007-1250	6.619
<i>Child Development</i>	0009-3920	4.718
<i>Clinical Psychology Review</i>	0272-7358	7.071
<i>Cognitive Psychology</i>	0010-0285	4.273
<i>Current Directions in Psychological Science</i>	0963-7214	3.929
<i>Depression and Anxiety</i>	1091-4269	4.184
<i>Development and Psychopathology</i>	0954-5794	4.397

<i>Developmental Science</i>	1363-755X	3.888
<i>Economic Geography</i>	0013-0095	3.975
<i>Emotion</i>	1528-3542	3.875
<i>Epidemiology</i>	1044-3983	5.566
<i>Evolutionary Anthropology</i>	1060-1538	3.594
<i>Global Environmental Change</i>	0959-3780	6.868
<i>Harvard Law Review</i>	0017-811X	3.948
<i>Health Affairs</i>	0278-2715	4.313
<i>Health Psychology</i>	0278-6133	3.873
<i>International Journal of Management Reviews</i>	1460-8545	3.581
<i>Journal of Abnormal Psychology</i>	0021-843X	4.857
<i>Journal of the American Academy of Child and Adolescent Psychiatry</i>	0890-8567	6.444
<i>Journal of the American Geriatrics Society</i>	0002-8614	3.737
<i>Journal of the American Medical Informatics Association</i>	1067-5027	3.609
<i>Journal of Applied Psychology</i>	0021-9010	4.308
<i>Journal of Child Psychology and Psychiatry</i>	0021-9630	4.281
<i>Journal of Clinical Psychiatry</i>	0160-6689	5.799
<i>Journal of Cognitive Neuroscience</i>	0898-929X	5.175
<i>Journal of Consulting and Clinical Psychology</i>	0022-006X	4.848
<i>Journal of Economic Literature</i>	0022-0515	9.243
<i>Journal of Economic Perspectives</i>	0895-3309	4.211
<i>Journal of Experimental Psychology: General</i>	0096-3445	3.986
<i>Journal of Finance</i>	0022-1082	4.218
<i>Journal of Financial Economics</i>	0304-405X	3.725
<i>Journal of Fluency Disorders</i>	0094-730X	4.05
<i>The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences</i>	1079-5006	4.598
<i>Journal of Human Evolution</i>	0047-2484	3.638
<i>Journal of Informetrics</i>	1751-1577	4.229
<i>Journal of Management</i>	0149-2063	4.595
<i>Journal of Management Studies</i>	0022-2380	4.255
<i>Journal of Marketing</i>	0022-2429	5.472
<i>Journal of Operations Management</i>	0272-6963	4.382
<i>Journal of Organizational Behavior</i>	0894-3796	3.854
<i>Journal of Personality and Social Psychology</i>	0022-3514	5.076
<i>Journal of Psychiatry & Neuroscience</i>	1180-4882	5.342
<i>Journal of Psychiatric Research</i>	0022-3956	4.664
<i>Kindheit und Entwicklung</i>	0942-5403	6

<i>Learning and Instruction</i>	0959-4752	3.732
<i>Milbank Quarterly</i>	0887-378X	5.62
<i>MIS Quarterly</i>	0276-7783	4.447
<i>Monographs of the Society for Research in Child Development</i>	0037-976X	5.5
<i>Neuropsychology Review</i>	1040-7308	6.618
<i>Neuropsychologia</i>	0028-3932	3.636
<i>Neuropsychology</i>	0894-4105	3.816
<i>Organization Science</i>	1047-7039	4.338
<i>Personality and Social Psychology Review</i>	1088-8683	6.071
<i>Perspectives on Psychological Science</i>	1745-6916	4.89
<i>Progress in Human Geography</i>	0309-1325	3.547
<i>Psychological Bulletin</i>	0033-2909	14.457
<i>Psychological Inquiry</i>	1047-840X	4.727
<i>Psychological Medicine</i>	0033-2917	6.159
<i>Psychological Methods</i>	1082-989X	4.449
<i>Psychological Review</i>	0033-295X	7.756
<i>Psychological Science</i>	0956-7976	4.431
<i>Psychosomatic Medicine</i>	0033-3174	3.968
<i>Psychotherapy and Psychosomatics</i>	0033-3190	6.284
<i>Quarterly Journal of Economics</i>	0033-5533	5.92
<i>Review of Financial Studies</i>	0893-9454	4.748
<i>Schizophrenia Research</i>	0920-9964	4.748
<i>Schizophrenia Bulletin</i>	0586-7614	8.8
<i>Social Cognitive and Affective Neuroscience</i>	1749-5016	6.132
<i>Stanford Law Review</i>	0038-9765	4.32
<i>Strategic Management Journal</i>	0143-2095	3.783
<i>Structural Equation Modeling</i>	1070-5511	4.71
<i>Transactions of the Institute of British Geographers</i>	0020-2754	3.536
<i>Trends in Cognitive Sciences</i>	1364-6613	12.586
<i>World Psychiatry</i>	1723-8617	6.233
<i>Yale Law Journal</i>	0044-0094	3.667