



# Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories



Youngseek Kim\*, Melissa Adler<sup>1</sup>

University of Kentucky, School of Library and Information Science, Little Library Building, Lexington, KY 40506, USA

## ARTICLE INFO

### Article history:

Available online 13 May 2015

### Keywords:

Data sharing  
Social scientist  
Individual motivation  
Institutional pressure  
Data repository

## ABSTRACT

The purpose of this study is to locate individual, institutional, and resource factors that influence data sharing behaviors among social scientists. Given the benefits to the social science disciplines in the advancement of scholarship, and the recent data sharing policy changes of funding agencies, it is necessary to determine the factors that support and impede data sharing behaviors. A research model was developed and validated based on the results of a survey of 361 social scientists. The model is informed by theory of planned behavior and institutional theory to map underlying individual motivations, institutional pressures, and availability of resources facilitating social scientists' data sharing. It was found that social scientists' data sharing behaviors are mainly driven by personal motivations (i.e., perceived career benefit and risk, perceived effort, and attitude toward data sharing) and perceived normative pressure. Funding agencies' pressure, journals' pressure, and availability of data repository were not found to be significant factors in influencing social scientists' data sharing. This research suggests that personal motivations and norm of data sharing currently support social scientists' data sharing; however, institutional pressures by funding agencies and journals and data repository need to be further encouraged to better facilitate social scientists' data sharing behaviors.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Raw data sets have become important "information currency" for scholarly communication (Davis & Vickery, 2007). Not only do data sets add value to traditional journal publications, they also increase transparency and facilitate high quality, continued research. Existing studies confirm, however, that data sharing in the social sciences is not widely practiced due to a variety of factors, including infrastructural and institutional barriers, ethical concerns, and personal reasons. The objective of the present study is to investigate the institutional and individual factors that influence social scientists' data sharing behaviors and to provide a model, based on institutional theory and the theory of planned behavior, for understanding and predicting behavior. Examining both individual motivation and institutional contexts provides a holistic view of data sharing practices across diverse social science disciplines and can inform policy-making decisions regarding the design and practices surrounding data archives. As data sharing is

not a norm, this study can help to identify ways to encourage and support data sharing in the social sciences.

Reasons for encouraging data sharing in the social sciences are many (Fienberg, 1994; King, 1995). Data sharing increases the transparency of quantitative analytic work, thereby lending more credibility to research findings, providing evidence to support analytic frameworks and decisions, and a source for researchers to consult when considering how to build upon existing studies. Having data openly available means that replication and verification is made immediately possible. Shared data allows testing different hypotheses and building better research studies. Furthermore, openly shared data facilitates participation from multiple perspectives, allowing access to the data for more disciplines and for researchers from different backgrounds. It reduces costs by avoiding the duplication of data collection efforts. Additionally, data made available through sharing contributes to the education of students. It is important to note that national scientific organizations and funding agencies have increasingly issued data archiving policies, and agencies including the National Science Foundation (NSF), National Institutes of Health (NIH), and the Institute for Museum and Library Services (IMLS) now require data sharing and management plans as part of grant applications. Given the benefits to the social science disciplines in the advancement of scholarship, and the recent data sharing policy changes of funding agencies, there is

\* Corresponding author. Tel.: +1 859 218 2295; fax: +1 859 257 4205.

E-mail addresses: [youngseek.kim@uky.edu](mailto:youngseek.kim@uky.edu) (Y. Kim), [melissa.adler@uky.edu](mailto:melissa.adler@uky.edu) (M. Adler).

<sup>1</sup> Tel.: +1 859 218 2294; fax: +1 859 257 4205.

a particularly pressing need to determine the factors that impede and support data sharing practices.

Data sharing has been defined somewhat differently in various studies. For the purpose of this research, data sharing is broadly defined as an individual scientist's behavior in providing their raw (or preprocessed) data of his/her published work to other scientists by making it accessible through central/local data repositories or by sending data via personal communication methods upon request.

The first sections of this paper present the literature relevant to the study and an overview of the theoretical frameworks that support our model for data sharing behavior. Section 4 describes the development of the model and hypotheses, and Section 5 provides a description of the research method, including how the survey data was collected and used. In Section 6 we present the results and analysis of the data, and Section 7 provides a discussion, which includes an account of qualitative data derived from participants' comments. The paper concludes by addressing limitations and considerations for further study.

## 2. Literature review

Researchers have examined a variety of dimensions related to data sharing. Across this research there is consensus that, although there is increasing awareness of the benefits of openly shared data, policies and standards are inconsistent across the social science disciplines and institutions, and data sharing is not a common practice among social scientists. Freese (2007) has argued that disciplines within the social sciences have different norms for data sharing, and suggests that much of the difference stems from a norm produced by the discipline's journals; whether they require authors to contribute data for publication to the journal website or simply allow authors to make the data available upon request. Bebeau and Monson (2011) provide an overview of data sharing statements set forth by professional associations in the fields of psychology, sociology, and education, and suggest that associations and scientific societies play an important role in supporting and encouraging data sharing. Pienta, Alter, and Lyle (2010), constructed a database comprised of administrative data from 40 years of social science research and found that "very few social science data collections are preserved and disseminated by an archive or institutional repository" (215). The same study found that informal data sharing is more common. At the global level, the International Federation of Data Organization (IFDO) issued a report in 2014, which provides the results of its survey of country-by-country information on funding agencies' policies regarding social science research data. The survey reveals a global recognition of the benefits of data sharing, but a wide variety in the existence and types of policies and their enforcement. It also found a lack or immaturity of infrastructure to implement or maintain data sharing policies in some "non-Western" regions (Kvalheim & Kvamme, 2014).

In the field of political science, Gary King defined the "replication standard" in 1995, arguing that political science research data should be shared: "Political science is a community enterprise; the community of empirical political scientists needs access to the body of data necessary to replicate existing studies to understand, evaluate, and especially build on this work. Unfortunately, the norms we have in place now do not encourage, or in some cases even permit, this aim." King continues to advocate data sharing but warns that the benefits of collecting and sharing data may be undermined by infrastructural weaknesses in managing the vast types and quantities of data (2011). He and others cited below have examined the merits, barriers, infrastructural design, and methods for data curation and sharing in the wider social sciences.

The actual data sharing rates in some fields of social sciences are even lower than what prior studies expected. Studies conducted

between 1962 and 2006 in the field of psychology consistently reveal data withholding behaviors. Wolins (1962) reported that, of the data requested from 37 authors who published articles in major APA (American Psychological Association) journals, only 9 authors responded with actual data sets, for a response rate of 24.3%. Similarly, Craig and Reese (1973) reported that 37.7% of authors (20 out of 53) provided either original data or a summary of data analysis in major APA journals. A 2006 study by Wicherts, Borsboom, Kats, and Molenaar (2006) shows that only 38 out of 141 authors (27.0%) responded with actual data sets upon request of research data for articles published in major APA journals. Even more striking are the recent findings of Andreoli-Versbach and Mueller-Langer (2014), who "provide evidence for the status quo in economics with respect to data sharing." They found that 8.81% of economics researchers share some data and only 2.05% fully share their data.

Although Freese stated in 2007 that sociology should be a leader in data sharing, Nicholson and Bennett (2011) find that, in practice, sociologists are not actually moving toward data sharing. They recommend a series of actions, which academic librarians can take to encourage data storage and sharing. They cite disagreements between Freese (2007) and Abbott (2007), who have differing opinions on the efficacy of data sharing. Whereas Freese "optimistically perceives a mandate for data access," Abbott raises concerns about the burden of evaluating sociological data sets on peer reviewers and privacy issues (p. 506).

Cliggett (2013) has discussed some of concerns specific to Anthropology and argues that anthropologists have an ethical and professional duty to share their primary data. She outlines a set of best practices for data preservation based on her own research in developing a system for digital archiving of qualitative ethnographic data.

Much of the existing data sharing research deals with the sciences more broadly, without placing an emphasis on the social science. Many of these studies focus on whether scientists allow or deny other researchers' access to their data (Campbell et al., 2002; McCain, 1991), and they identify diverse factors influencing scientists' data sharing behaviors including individual factors, institutional factors, and resource factors. With regard to individual factors, prior studies located perceived benefits (Kim, 2007; Kling & Spector, 2003), reciprocal benefit (Zimmerman, 2007), and perceived efforts (Campbell et al., 2002; Louis, Jones, & Campbell, 2002; Tenopir et al., 2011). Also, a number of studies reported perceived risks involved in data sharing as a main barrier for data sharing, and those risk factors include losing publication opportunities (Campbell et al., 2002; Savage & Vickers, 2009), losing commercialization opportunities (Blumenthal et al., 2006; Tenopir et al., 2011), misuse (Borgman, Wallis, & Enyedy, 2007; Cragin, Palmer, Carlson, & Witt, 2010), and privacy issues (Borgman, 2009; Savage & Vickers, 2009).

With regard to institutional factors, prior studies on data sharing in the sciences (broadly conceived) have investigated funding requirements (McCullough, McGeary, & Harrison, 2008; Piwowar, 2011), journal policies (Piwowar & Chapman, 2008; Savage & Vickers, 2009), and contracts with industry sponsors (Campbell & Bendavid, 2003; Louis et al., 2002). Scholars found that scientists who received funding from government agencies (e.g. NSF and NIH) were more likely to share their data with others (McCullough et al., 2008; Piwowar, 2011). However, Campbell and Bendavid (2003) found that government agencies sometimes provide scientists with funding under strict restrictive policies about data sharing if a national security issue is at stake. In terms of journal policy, Piwowar and Chapman (2008) found that there is a positive correlation between the strength of journals' data sharing policies and the rate at which scientists deposit microarray data in a public database. Another study by Savage and Vickers (2009), however, investigated whether the authors whose articles were published

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات