



## Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories



Ayoung Yoon<sup>a</sup>, Youngseek Kim<sup>b,\*</sup>

<sup>a</sup> Indiana University-Purdue University, School of Informatics and Computing, Department of Library and Information Science, 535 W. Michigan St. IT563, Indianapolis, IN 46202, United States

<sup>b</sup> University of Kentucky, School of Information Science, 331 Little Library Building, Lexington, KY 40506, United States

### A B S T R A C T

Many disciplines within the social sciences have a dynamic culture of sharing and reusing data. Because social science data differ from data in the hard sciences, it is necessary to explicitly examine social science data reuse. This study explores the data reuse behaviors of social scientists in order to better understand both the factors that influence those social scientists' intentions to reuse data and the extent to which those factors influence actual data reuse. Using an integrated theoretical model developed from the theory of planned behavior (TPB) and the technology acceptance model (TAM), this study provides a broad explanation of the relationships among factors influencing social scientists' data reuse. A total of 292 survey responses were analyzed using structural equation modeling. Findings suggest that social scientists' data reuse intentions are directly influenced by the subjective norm of data reuse, attitudes toward data reuse, and perceived effort involved in data reuse. Attitude toward data reuse mediated social scientists' intentions to reuse data, leading to the indirect influence of the perceived usefulness and perceived concern of data reuse, as well as the indirect influence of the subjective norm of data reuse. Finally, the availability of a data repository indirectly influenced social scientists' intentions to reuse data by reducing the perceived effort involved.

### 1. Introduction

There is a long tradition of sharing and reusing data in the social sciences. Hedrick (1988) argues that data sharing has been a concern for researchers since the late 1970s. However, while there were (and are) difference within disciplines, discussions about the value and sharing of social science data began in the early 1960s (Clubb, Austin, Gedda, & Traugott, 1985). For decades, the topic has intrigued researchers working with large-scale survey data, archivists at institutional repositories, and individuals who were frustrated with unsuccessful attempts to obtain other researchers' data. Fear (2013) asserts that this tradition of sharing and reusing data in the social sciences is due to the nature of social research, which often requires large amounts of unique data collected over time.

While there is no agreed upon formal definition of “social science data”, the term has been generally understood to mean “numeric files originating from social research methodologies or administrative records, from which statistics are produced” (Inter-university Consortium for Political and Social Research [ICPSR], 2016). As implied by this definition, quantitative data have been the dominant form of data in

social science, and Fear (2013) states that reuse of such data from repositories is the most common type of data reuse in social science. Other types of data have been also generated and reused in social science; for instance, qualitative data reuse is an established practice in some social science disciplines (Yoon, 2014b) and discussions of qualitative data sharing and reuse have emerged in journals such as *Forum: Qualitative Social Research* (Bergman & Eberle, 2005) and *IASSIST Quarterly* (Rasmussen, 2010).

While the social sciences, broadly speaking, have had a dynamic culture of sharing and reusing data, much of the research on data reuse in recent years has focused primarily on the life and physical sciences. Social science data differ from the data from lab-based or other life and physical science research. Social science data typically involve observations about human subjects and unstructured formats (e.g., interview transcripts, observation notes, and survey data). The data practices of social science research are arguably different as well; because they involve human subjects, they are usually regulated by institutional review boards. Understanding and interpreting the unstructured data collected in the social sciences often requires detailed contextual information. Given the breadth and importance of the

\* Corresponding author.

E-mail addresses: [ayyoon@iupui.edu](mailto:ayyoon@iupui.edu) (A. Yoon), [youngseek.kim@uky.edu](mailto:youngseek.kim@uky.edu) (Y. Kim).

differences of social science data, its reuse cannot be understood without studying it explicitly and exclusively from life and physical science data.

## 2. Problem statement

Several recent studies have investigated data reuse practices and behaviors in the social sciences (Daniels, Faniel, Fear, & Yakel, 2012; Faniel, Kriesberg, & Yakel, 2012; Faniel, Kriesberg, & Yakel, 2016; Niu, 2009; Yoon, 2014b, 2016, 2017) as part of a broader drive to understand data practices within the social sciences. While these studies have captured some of the contexts and characteristics of social science data reuse by focusing on specific aspects of data reuse practices, which may overlap with other disciplinary contexts, fewer studies have used theoretical approaches or models to explain social scientists' data reuse behaviors.

A theoretical model is often used to explain the meaning, nature, and challenges associated with the phenomena of interest, and it helps scholars to understand these phenomena more effectively. The lack of a theoretical model in the data reuse research leaves a significant gap in our understanding of the way disciplinary, organizational, and individual characteristics interact to encourage or discourage data reuse. This study advances a theoretical model of social scientists' data reuse behaviors. Specifically, this study explores the factors that influence social scientists' intentions to reuse data and the extent to which those factors influence actual data reuse. The theoretical model provides a broad explanation of the relationships that exist between factors that influence social scientists' data reuse. The conceptual underpinnings of this study will provide a new perspective for understanding data reuse behaviors, and will contribute to both theory and practice.

Although data reuse is important for many natural science and engineering disciplines, data reuse is becoming increasingly significant in social sciences in the context of data-intensive research as researchers reuse shared data sets along newly collected data sets. A better understanding of data reuse in the social sciences can help social scientists to compare their studies to existing ones and conduct more advanced studies based on shared and accumulated data sets. This study can also offer valuable insights for academic libraries seeking to develop or improve data stewardship services by taking into account the diverse factors affecting social scientists' data reuse behaviors.

## 3. Literature review

Many researchers discuss the social and individual benefits of data reuse. Data reuse expands research possibilities and saves on data (re) collection costs (Borgman, 2012). Yoon (2015), based on empirical research targeting social scientists, reported the perceived benefits of and motivations for using existing data, which included the data reusers' awareness of the usefulness of secondary data, the cost-effectiveness of reusing data, the ability to use large sample data, and the expediency of reusing data for training and education. Although there are some common benefits and motivations reported by others for all data reuse, Yoon (2015) found that the reuse of data gathered from large samples can be particularly helpful in verifying and generalizing prior findings in quantitative social science research. Curty (2015) also found that social scientists' data reuse intentions are mainly affected by perceived benefit involved in data reuse.

Despite the potential benefits of data reuse, many social scientists still have concerns, and it is known that they have more concerns about qualitative data reuse than about other types of data reuse. Bishop (2009) reported that qualitative researchers expressed concerns about potential ethical violations, since qualitative research involves direct interaction with human subjects. In addition, although the possibility of misinterpretation is a concern with all data reuse, qualitative researchers are more concerned with the nature of their data in general, because "knowledge about qualitative data is highly contextual and

experience-dependent" (Niu & Hedstrom, 2008, p. 7). Reused data can also be perceived as less valuable (Goodwin, 2012; Martin, 1995), and the qualitative researchers in Yoon's (2014b) study faced challenges publishing their work which reused existing data; this too raised concerns about reusing data. Curty (2015) also reported that perceived risk involved in data reuse significantly affected social scientists' data reuse intentions.

Discovering relevant data may be challenging for scientists across disciplines (e.g., Faniel & Jacobsen, 2010; Zimmerman, 2008), but it is especially difficult for social scientists because data are distributed among various sources and systems (Yoon, 2015). Easy access to data was one of the most influential factors in determining social scientists' satisfaction with data reuse (Faniel, Kriesberg, & Yakel, 2016). Data repositories have a long history in social science, and they are known to support easy access to and reuse of available data through value-adding activities (Daniels, Faniel, Fear, & Yakel, 2012; Yoon, 2014a). However, social scientists searched for more data than was deposited in the repositories. In addition, Curty (2015) found that social scientists' data reuse is influenced by facilitating conditions such as documentation, repository, support, and training.

Even when data reusers can find sufficient, seemingly suitable data, data reuse can still pose challenges. Data reusers need to assess data before reusing it because they are usually unfamiliar with the details of the data. Reusers assess data for a good fit for the purpose of their study (Faniel, Kansa, Kansa, Barrera-Gomez, & Yakel, 2013), for data quality (Cragin & Shankar, 2006; Van House, 2002), or generally for reusability (Faniel & Jacobsen, 2010). Social scientists are also concerned with choosing good quality, trustworthy data and avoiding data with errors (Yoon, 2014a, 2016, 2017). Assessing data for each of these qualities requires different criteria; some important assessment factors which have been identified include data producers' ability to generate trustworthy data, other reusers' positive experiences using the data, and soundness of methodology used to produce data (Faniel & Jacobsen, 2010; Faniel, Kansa, Kansa, Barrera-Gomez, & Yakel, 2013; Yoon, 2017; Zimmerman, 2008).

A particular challenge arises from the fact that reusers have not participated in the initial study design and data collection process; thus, it can be difficult for them to understand the data. Issues arising from the contextual nature of data and the fundamental challenges of transferring contextual information to data reusers exist across disciplines (e.g., Berg & Goorman, 1999; Cragin & Shankar, 2006; Faniel et al., 2013; Jirotko et al., 2005). Documentation can play an important role in transferring contextual information and supporting data reuse, but reusers reported different experiences working with documentation and of its usefulness (Borgman, 2007; Faniel et al., 2013). Markus (2001) differentiates documentation for oneself, similar others, and dissimilar others and argues that the level of detail and types of contextual information included in the documentation should be different depending on the intended users. According to Niu (2009), documentation for quantitative data in social science tends to be better than that of other kinds of data.

Several studies have demonstrated that human interactions also play an important role in data reuse. Data reusers often search for additional information when documentation is insufficient, consulting various sources, including data producers and experts (Birnholtz & Bietz, 2003; Bishop, 2009; Faniel et al., 2013; Markus, 2001; McCall & Appelbaum, 1991). Yoon (2017) found that social science data reusers also sought external help from data reuser groups, repository staff, and data producers when they encountered problems. Faniel, Kriesberg, and Yakel (2012) reported that human scaffolding, particularly the use of faculty advisors, was an effective technique for novice social science data reusers to manage complex issues that arose during data reuse.

While these studies contribute to the understanding of data reuse practices in the social sciences, explicating relevant factors in data reuse and explaining social science data reuse as compared to other

Download English Version:

<https://daneshyari.com/en/article/5123876>

Download Persian Version:

<https://daneshyari.com/article/5123876>

[Daneshyari.com](https://daneshyari.com)