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

Ayoung Yoon\textsuperscript{a}, Youngseek Kim\textsuperscript{b,⁎}

\textsuperscript{a} 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

\textsuperscript{b} 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
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, 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., Borg & Goorman, 1999; Cragin & Shankar, 2006; Faniel et al., 2013; Jirotka 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...
disciplines, only a few studies try to explain or understand data reusers' behaviors from theoretical perspectives or using a theoretical model (e.g., Yoon, 2017). This study develops an integrated theoretical framework to explore social scientists' data reuse behaviors.

4. Conceptual development

4.1. Theory of planned behavior (TPB) and technology acceptance model (TAM)

This study employs an integrated theoretical framework combining the theory of planned behavior (TPB) and the technology acceptance model (TAM). TPB is a social psychology theory linking individuals' behaviors to their attitude toward the behavior, subjective norms, and perceived behavioral controls involved in the behaviors (Ajzen, 1991; Fishbein & Ajzen, 1975). In TPB, attitude, subjective norm, and behavioral control are determined by an individual's attitudinal, normative, and control beliefs (Ajzen, 1991; Fishbein & Ajzen, 1975). Those attitudinal, normative, and control beliefs are formed by a person's fundamental thoughts or views on the results of a certain behavior.

TPB was used in this study to understand social scientists' data reuse behaviors by considering (1) their attitudes toward data reuse, i.e., their overall evaluation of reusing others' data, (2) social scientists' subjective norms of data reuse, i.e., the community expectations of social scientists about data reuse, and (3) perceived behavioral controls (or, resource facilitating conditions) such as data repositories or institutional supports. TPB uses intentions as a proxy for actual behaviors; TPB explains how attitudes, subjective norms, and perceived behavioral controls all influence individuals' intentions to engage in a certain behavior.

Although TPB provides a fundamental theoretical framework for understanding how human behavior is influenced by one's attitude, subjective norm, and behavioral control factors, TPB does not reveal what specific beliefs influence that attitude, norm, or behavior control. Therefore, this study also used TAM to better account for social scientists' perceptions of data reuse. Since the research construct of "attitude toward data reuse" may be influenced by diverse attitudinal beliefs, this research employed TAM to explain specific perceptions toward data reuse. TAM considers utility and effort expectancies, such as perceived usefulness and perceived ease of use, to explain people's intentions to adoption a technology. Integrating TAM with TPB makes sense, as TAM can provide two important research constructs, perceived usefulness of data use and perceived ease of use (e.g., perceived ease of use), involved in data reuse. The TAM construct of "perceived ease of use" was adapted to "perceived effort" to better deliver the idea of effort expectancy involved in data reuse. The utility and effort expectancies captured by TAM are social scientists' perception that data reuse will benefit them, and be worth their effort. In addition to perceived usefulness and perceived effort, this research integrated the perceived concern involved in data reuse to better explain social scientists' attitude toward data reuse. Perceived concern comprises potential risks involved in data reuse. Social scientists' may not want to reuse other scientists' data if they might infringe upon copyright or not be able to publish their research based on the reused data.

Given the complexity of data reuse intention and behavior, neither theory is sufficient on its own. The integration of TPB and TAM provides a necessary framework by (1) providing a fundamental theoretical model considering attitude, subjective norm, and behavioral control altogether (TPB), and (2) suggesting specific attitudinal beliefs which mediate social scientists' attitudes toward data reuse, thereby changing their behavioral intentions.

4.2. Research model and hypotheses development

Based on the integrated theoretical framework above, a research model for social scientists' data reuse was designed. Perceptions were tested, including perceived usefulness, concern, effort involved in data reuse and the influences of these perceptions on social scientists' attitudes toward data reuse. The "subjective norm of data reuse" was also included as part of TPB's subjective norm construct. In TPB, subjective norm is believed to influence both people's behavioral intentions and their attitudes toward a certain behavior (Ajzen, 1991; Ajzen & Fishbein, 2005). The existence of a data repository was also included as an external behavioral control factor which, under TPB, would have a direct influence on people's behavioral intentions and effort expectancy in data reuse.

4.2.1. Perceived usefulness

"Perceived usefulness" refers to the degree to which social scientists believe they would benefit by reusing other researchers' data. In TAM, perceived usefulness is one of the constructs for determining system use; it is understood as the degree to which a person believes using a system will increase relevant job performance (Davis, 1989). In data reuse literature, usefulness is mostly relevant to the benefits that researchers find from reusing existing data, including the increase in research performance and effectiveness. Researchers' decision to reuse data was primarily based on the perceived benefits or the perception that the data met the researchers' needs (Niu, 2009; Yoon, 2015). Pienta, Alter, and Lyle (2010) also considered research productivity as one of the drivers of data reuse. Data reusers are aware that reusing data can increase their research productivity (e.g., increased number of publications, reduced time for data collection). Thus, the perceived usefulness of data reuse should improve social scientists' attitudes toward data reuse.

H1. Perceived usefulness positively affects a social scientist's attitude toward data reuse.

4.2.2. Perceived concern

"Perceived concern" refers to the degree to which social scientists believe data reuse would involve possible risks, such as fewer publication opportunities, misinterpretation of data, or copyright infringement. In the context of social science data reuse, researchers expressed concern about data reuse because they had experienced challenges when trying to publish articles reusing existing data (Yoon, 2014b). As Martin (1995) and Goodwin (2012) suggest, reused data might be perceived as less acceptable or less valuable in some social science disciplines. The possibility of misrepresentation arising from missing information or missing context was also a major concern, not only for data producers but also for data reusers, particularly qualitative researchers (Niu & Hedstrom, 2008; Yoon, 2014b). Perceived concern comprises potential risks involved in data reuse, and TAM was extended by later studies including a perceived risk construct (Featherman & Pavlou, 2003; Wu & Wang, 2005). Given these potential problems, the perceived concern involved in data reuse could lead social scientists to hold negative attitudes about reusing other scientists' data.

H2. Perceived concern negatively affects a social scientist's attitude toward data reuse.

4.2.3. Perceived effort

"Perceived effort" refers to the degree to which social scientists believe data reuse would require time and energy in order to acquire other scientists' data and to process that data. In TAM, Davis (1989) used "ease of use" as a construct, referring to the degree to which a person believes using a system will be effort-free. "Perceived effort", rather than "perceived ease of use," emphasizes the effort expectancy involved in data reuse as compared to the effortlessness or ease of data reuse that a "perceived ease of use" construct would imply (Venkatesh, Morris, Davis, & Davis, 2003). Data reuse literature suggests that when data reusers spend significant time actually using
existing data (Rolland & Lee, 2013; Zimmerman, 2008), and less time acquiring and processing data, their level of satisfaction with data reuse improves (Faniel et al., 2016). Thus, it is likely that a greater perceived effort to reuse data would negatively influence social scientists’ attitudes toward data reuse and their intention to reuse other scientists’ data.

H3. Perceived effort negatively affects a social scientist’s attitude toward data reuse.

H4. Perceived effort negatively affects a social scientist’s intention to reuse other scientists’ data.

4.2.4. Subjective norm of data reuse

The “subjective norm of data reuse” refers to the degree to which social scientists consider data reuse a prevalent research practice in their research communities. Not many data reuse studies have examined the impact of disciplinary norms on data reuse, but previous studies that focused on a specific discipline or across different disciplines suggested that reuse practices were varied and discipline-specific (e.g., Birnholtz & Bietz, 2003; Carlson & Anderson, 2007; Faniel et al., 2013; Rolland & Lee, 2013). This may imply that the norm in a discipline would influence reuse behaviors. In addition, Yoon’s (2014b) study indicated that the lack of data reuse norms in certain social science disciplines negatively influences researchers’ data reuse behaviors. Thus, the subjective norm of data reuse would positively influence social scientists’ attitudes toward data reuse and encourage their intentions to reuse other scientists’ data.

H5. Subjective norm of data reuse positively affects a social scientist’s attitude toward data reuse.

H6. Subjective norm of data reuse positively affects a social scientist’s intention to reuse other scientists’ data.

4.2.5. Attitude toward data reuse

“Attitude toward data reuse” refers to the degree to which social scientists believe data reuse is good. The “attitude” in TPB is usually understood as a summary evaluation of a certain behavior, and TPB shows that the attitude toward a certain behavior strongly explains the intention to conduct the behavior (Ajzen, 1991; Ajzen & Fishbein, 2005; Fishbein & Ajzen, 1975). A good number of studies of knowledge sharing have found that positive attitudes toward sharing knowledge lead to positive intentions to share knowledge (Cho, Chen, & Chung, 2010; He & Wei, 2009; Hsu & Chiu, 2004). Thus, a positive attitude toward data reuse would encourage social scientists’ intentions to reuse other scientists’ data.

H7. Positive attitude toward data reuse positively affects a social scientist’s intention to reuse other scientists’ data.

4.2.6. Availability of data repository

The availability of a data repository is an important factor influencing social scientists’ data reuse. The resource-facilitating conditions found in data repositories can be considered external behavioral controls in TPB (Ajzen, 1991; Ajzen & Fishbein, 2005). Data repositories can reduce perceived effort, enhance positive attitudes, and influence the actual behaviors of social scientists regarding data reuse. In the social science context, data repositories play an important role in data sharing and reuse, although data reuse also still occurs through person-to-person exchange and interaction (e.g., Faniel & Jacobsen, 2010; Yoon, 2014a). Social science data repositories contribute to data reuse, not just through easy access to data, but also by providing value-added services which ensure current and future use of that data, such as managing provenance, correcting errors, and providing supporting documentation, all of which reduce the effort expectancy required for data reuse (Daniels et al., 2012; Fear & Donaldson, 2012). Data repositories also help data reusers trust data because of the repositories’ function, reputation, and structure, which encourages researchers to reuse data from repositories (Yakel, Faniel, Kriesberg, & Yoon, 2013; Yoon, 2014a). Thus, the availability of a data repository would reduce social scientists’ perceived effort and positively influence their intentions to reuse other scientists’ data.

H8. The availability of a data repository negatively affects a social scientist’s perceived effort.

H9. The availability of a data repository positively affects a social scientist’s intention to reuse other scientists’ data.

4.2.7. Intention to reuse data

Data sharing and reuse are not yet well-established research practices in most social science disciplines; because of that limitation, this research measured social scientists’ “intentions” to reuse other researchers’ data rather than their actual data reuse behaviors. However, while this choice is a practical one, it is nonetheless firmly grounded. Prior knowledge sharing studies, measures of people’s intention to engage in a particular behavior serve as a proxy for the actual behavior (Bock, Zmud, Kim, & Lee, 2005; Hsu & Lin, 2008; Kuo & Young, 2008; Lin, 2006). Furthermore, prior studies employing TPB as their theoretical framework support the strong correlation between intentions and actual behaviors (Chen & Chen, 2009; Kuo & Young, 2008; Tsai & Cheng, 2010); therefore, this research is well founded in its use of “intentions” to reuse data as its main outcome variable, for both methodological and theoretical reasons.

Fig. 1 places the hypotheses in the context of the research model incorporating the constructs described above.

5. Research method

Survey methodology was used to evaluate social scientists’ data reuse based upon the research constructs in the model. Surveys are
useful in eliciting people's perceptions of particular behaviors and their intentions to perform those behaviors (Ajzen & Fishbein, 2005; Creswell, 2008).

5.1. Population and sampling

The Community of Science (COS) Scholars database was the major recruitment source. COS includes social science faculty members, post-doctoral researchers, and graduate student researchers in U.S. academic institutions; 96,121 social scientists from the United States were registered in the database (as of September 4th, 2015). 2193 social science scholars were randomly selected and invited to participate. Since structural equation modeling (SEM) was used for data analysis method, a minimum of 200 responses were required to appropriately evaluate the model (Chin, 1998; Fornell & Larcker, 1981).

Compared to prior data reuse studies (Faniel et al., 2016; Yoon, 2014b), this research considered an inclusive range of social science researchers in terms of their data reuse experience (e.g., experienced data reusers, potential data reusers, and even those who never want to reuse data). The survey is also designed to incorporate the wide variety of social science researchers who may have different perceptions (e.g., subjective norm of data reuse) and intentions (e.g. potential data reusers vs those who may not intend to reuse data). Measuring data reuse intentions as the proxy of actual reuse behaviors allowed the researchers to identify important factors which might not have surfaced had the study been limited to subjects with actual data reuse experience.

5.2. Measurement of constructs

The survey questionnaire was designed to measure the research constructs. A total of 21 measurement items for 7 research constructs were adapted from prior studies: perceived usefulness (Davis, 1989), perceived concern (Featherman & Pavlou, 2003), perceived effort (Davis, 1989), attitude toward data reuse (Ajzen & Fishbein, 2005), subjective norm of data reuse (Ajzen & Fishbein, 2005; Scott, 2001), and availability of data repository (Venkatesh, Morris, Davis, & Davis, 2003). In addition, the measurement items for intentions to reuse other scientists’ data were adopted from Ajzen and Fishbein’s (2005) study. 5-point Likert scales ranging from “strongly agree” to “strongly disagree” were used for the most of survey questions.

5.3. Survey administration

Before distributing the survey, permission was obtained from ProQuest Pivot to use the COS Scholars database. The online survey included a brief description of the study, the study purpose, and the measurement items for each construct (Appendix A). The survey consisted of 27 items, including some demographic questions. On October 5, 2015, the first invitation emails were sent to the 2193 social scientists who were randomly selected from the COS Scholars database. Only one reminder was sent, on November 10, 2015, before the survey was closed on November 30, 2015. Of the 2193 messages sent, 234 messages were returned because of invalid email addresses or spam filters in email servers. A total of 1959 messages were delivered to potential survey participants, and 292 valid responses were received with < 5% of missing values. The response rate was 14.91%. These 292 responses were used for the final data analysis. Only 9 responses out of 292 responses had < 5% of missing values, and mean replacement method was used to treat these missing data in the final data analysis.

5.4. Demographics of participants

The survey participants were mixed in gender (male: 58.9%; female: 39.0%) and age (from 20s to 60s). Caucasians (80.1%) were the most dominant group of respondents. Most respondents were researchers with a Ph.D. degree (91.1%); and more than half of respondents were also tenured or tenure track faculty (62.3%). Their positions were varied but 77.1% held some rank of professorship (from assistant, associate, and full to emeritus). Table 1 summarizes the demographic data.

The survey participants were from diverse social science disciplines (Table 2). Almost a third (30.1%) were in areas of psychology, including clinical psychology (5.1%), non-clinical psychology (5.1%), combined psychology (3.1%), and other areas (16.80%). Sociology (13.0%), anthropology (12.3%), and political science (9.6%) were next most frequently represented.

6. Data analysis and results

Partial least square (PLS) based SEM technique was the primary data analysis method using SmartPLS 2.0 (Ringle, Wende, & Will, 2005). Following the two-stage approach proposed by Anderson and Gerbing (1988), we first tested a measurement model was first tested to examine the reliability and validity of each research construct, then, the structural model was followed to evaluate the research model and hypotheses.

---

Table 1
Demographics.

<table>
<thead>
<tr>
<th>Demographic category</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>172</td>
<td>58.9</td>
</tr>
<tr>
<td>Female</td>
<td>114</td>
<td>39.0</td>
</tr>
<tr>
<td>Missing</td>
<td>6</td>
<td>2.1</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>21</td>
<td>7.2</td>
</tr>
<tr>
<td>35-44</td>
<td>61</td>
<td>20.9</td>
</tr>
<tr>
<td>45-54</td>
<td>78</td>
<td>26.7</td>
</tr>
<tr>
<td>55-64</td>
<td>76</td>
<td>26.0</td>
</tr>
<tr>
<td>65+</td>
<td>55</td>
<td>18.8</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Ethnic</td>
<td>224</td>
<td>78.9</td>
</tr>
<tr>
<td>Caucasian</td>
<td>204</td>
<td>80.1</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>18</td>
<td>6.2</td>
</tr>
<tr>
<td>Other/multi-racial</td>
<td>14</td>
<td>4.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>12</td>
<td>4.1</td>
</tr>
<tr>
<td>Black/African-American</td>
<td>5</td>
<td>1.7</td>
</tr>
<tr>
<td>Native American/Alaska Native</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>Missing</td>
<td>6</td>
<td>2.1</td>
</tr>
<tr>
<td>Education</td>
<td>266</td>
<td></td>
</tr>
<tr>
<td>PhD/doctoral degree</td>
<td></td>
<td>91.1</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>20</td>
<td>6.8</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>Status</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Tenured</td>
<td>163</td>
<td>61.7</td>
</tr>
<tr>
<td>Not on tenure track</td>
<td>30</td>
<td>10.8</td>
</tr>
<tr>
<td>Retired</td>
<td>27</td>
<td>9.4</td>
</tr>
<tr>
<td>On tenure track</td>
<td>5</td>
<td>1.7</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>Position</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Full professor</td>
<td>81</td>
<td>36.2</td>
</tr>
<tr>
<td>Associate professor</td>
<td>76</td>
<td>30.3</td>
</tr>
<tr>
<td>Professor emeritus</td>
<td>28</td>
<td>9.8</td>
</tr>
<tr>
<td>Assistant professor</td>
<td>23</td>
<td>7.7</td>
</tr>
<tr>
<td>Researcher</td>
<td>13</td>
<td>4.5</td>
</tr>
<tr>
<td>Lecturer/instructor</td>
<td>11</td>
<td>3.7</td>
</tr>
<tr>
<td>Graduate student</td>
<td>11</td>
<td>3.7</td>
</tr>
<tr>
<td>Post-doctoral fellow</td>
<td>6</td>
<td>2.4</td>
</tr>
<tr>
<td>Professor of practice</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
<td>8.2</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Total</td>
<td>292</td>
<td>100</td>
</tr>
</tbody>
</table>

---

1 Both survey data and instrument have been made publicly available via Open ICPSR and can be accessed at [http://www.openicpsr.org/openicpsr/project/100404/view](http://www.openicpsr.org/openicpsr/project/100404/view).
indicating reliable convergent and discriminant validity. The measurement model evaluation showed that the measurements for each research construct are reliable and valid for further analysis using structural model evaluation.

The validity of the instrument, including convergent and discriminant validity, was tested by comparing the correlations between one item and the others, and the square root of its AVE. The square roots of AVEs have higher values (convergent validity) than the correlations between the items comprising different constructs (discriminant validity). The square roots of AVEs between items of the same construct ranged from 0.83 (perceived usefulness) to 0.97 (intention to reuse data), which are greater than the inter-item correlation coefficients between items of the different constructs (ranging from −0.04 to 0.75).

### 6.2. Structural model

Since the research constructs showed acceptable convergent and discriminant validity, data analysis proceeded and the structural model was estimated using the PLS-SEM technique. The results show that attitudinal, normative, and resource factors all strongly influence social scientists’ intentions to reuse others’ data. In terms of attitudinal beliefs, perceived usefulness was found to have a significant positive influence on attitude toward data reuse ($\beta = 0.625, p < 0.001$), and perceived concern was found to have a significant negative influence on attitude toward data reuse ($\beta = −0.215, p < 0.001$). However, perceived effort was not found to have any significant influence on social scientists’ attitudes toward data reuse ($\beta = 0.072, p > 0.05$). Along with attitudinal belief factors, findings also indicated that social scientists’ subjective norms of data reuse had a significant positive influence on their attitudes toward data reuse ($\beta = 0.096, p < 0.05$). The perceived usefulness, concern involved in data reuse, and subjective norm of data reuse factors accounted for 60.3% of total variance in attitude toward data reuse ($R^2 = 0.603$).

Social scientists’ subjective norms of data reuse showed significant, positive relationships with both attitudes toward data reuse ($\beta = 0.096, p < 0.05$) and intention to reuse data ($\beta = 0.169, p < 0.01$). The availability of a data repository was found to have a moderate negative influence on the perceived effort involved in data reuse ($\beta = −0.176, p < 0.01$); however, it was not found to have any significant direct influence on intention to reuse data ($\beta = 0.091, p > 0.05$). Perceived effort, as a behavioral control variable, was found to have a significant negative influence on intention to reuse data ($\beta = −0.153, p < 0.001$) directly. Lastly, attitude toward data reuse was found to have a significant positive influence on intention to reuse data ($\beta = 0.467, p < 0.001$). The subjective norm of data reuse, attitude toward data reuse, and perceived effort explain 38.4% of total variance in social scientists’ intentions to reuse data ($R^2 = 0.384$). Fig. 2 shows the results of hypothesis testing based on the model, and Table 5 summarizes hypothesis testing results.
reuse intentions, in accordance with TPB.

and the availability of a data repository, a
toward data reuse, along with subjective norms of data reuse
beliefs (i.e., perceived usefulness and concern) in
fl
7. Discussion

The integrated theoretical framework explains which attitudinal
beliefs (i.e., perceived usefulness and concern) influence social scientis-
ts’ attitudes toward data reuse based on TAM, and then how those
attitudes toward data reuse, along with subjective norms of data reuse
and the availability of a data repository, affect social scientists’ data
reuse intentions, in accordance with TPB.

Attitudes toward data reuse were influenced by perceived usefulness
and perceived concern. Social scientists consider reusing others’
data because they perceive that doing so would increase their research
performance and productivity. However, they are hesitant to reuse
others’ data when they think doing so could potentially cause problems,
such as misrepresentation of data, copyright infringement, and/or
fewer publication opportunities. In addition, social scientists develop
positive attitudes toward data reuse if they believe that their com-
mu-nities or disciplines have strong norms of data reuse. These findings
demonstrate the significance of informing and educating social scientis-
ts about the potential and bene-
fi-cs of data while mitigating possible
concerns. Since having a positive attitude toward data reuse influences
intention to reuse data, educating and informing them would be an
important first step toward actual data reuse.

It is worth noting that the established community/disciplinary
norms of data reuse were found to have significant positive impacts on
both attitudes toward data reuse and intentions to reuse data in this
study. When social scientists perceive an expectation of data reuse and
acknowledgement by their communities and disciplines that data reuse
is common and acceptable, they are more likely to have strong norms of
data reuse and to have positive attitudes toward data reuse. This sug-
gests that social science communities need a better and stronger sub-
jective norm of data reuse to lead to more active data reuse behavior
among researchers. Not many research studies have reported on what
the norm for data reuse in each field of social science is yet, but it is true
that some fields have a more active culture of data reuse, where others
do not. Even within a particular field, the subjective norm of data reuse
may vary depending on the types of research done. These differing
norms might well be reflected in differing levels of actual data reuse.
Further exploring the data reuse behaviors in a variety of dis-
ciplines—both with and without strong norms of data reuse—will help
us to better understand the impact of norms in data reuse behaviors and
lead to a better understanding of how a community can create a data
reuse culture.

Perceived effort of data reuse has a strong negative influence on
social scientists’ data reuse intentions. In addition, the availability of
a data repository was an important resource factor found to have a sig-
nificant negative relationship with the perceived effort involved in data
reuse. Although the availability of a data repository was not found to
have a significant relationship with data reuse intentions, it influences
data reuse intention indirectly through its impact upon perceived effort.
This finding suggests that the availability of a data repository can re-
duce the effort expectancy involved in data reuse, and lower effort
expectancy can increase social scientists’ intentions to reuse others’
data. When social scientists believe reusing others’ data requires too
much effort, they are less likely to reuse data. It is important to reduce
their effort expectancy by providing resources and support, such as data
repositories, to encourage data reuse.

7.1. Theoretical implications

This research has significant theoretical and practical implications.
While previous studies on data reuse have empirically investigated re-
searchers’ experiences, perceptions, and attitudes regarding data reuse,
few have employed a theoretical approach to explain behaviors and to
investigate the relationships among various factors that influence data

Table 5
Summary of hypothesis testing results.

<table>
<thead>
<tr>
<th>Hs Statements</th>
<th>Result</th>
<th>Beta (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Perceived usefulness would positively affects a social scientist’s attitude toward data reuse</td>
<td>Supported</td>
<td>0.625**</td>
</tr>
<tr>
<td>H2 Perceived concern would negatively affects a social scientist’s attitude toward data reuse</td>
<td>Supported</td>
<td>−0.215**</td>
</tr>
<tr>
<td>H3 Perceived effort would negatively affects a social scientist’s attitude toward data reuse</td>
<td>NOT Supported</td>
<td>0.072</td>
</tr>
<tr>
<td>H4 Perceived effort would negatively affects a social scientist’s intention to reuse other scientists’ data</td>
<td>Supported</td>
<td>−0.153***</td>
</tr>
<tr>
<td>H5 Subjective norm would positively affects a social scientist’s attitude toward data reuse</td>
<td>Supported</td>
<td>0.096</td>
</tr>
<tr>
<td>H6 Subjective norm would positively affects a social scientist’s intention to reuse other scientists’ data</td>
<td>Supported</td>
<td>0.169</td>
</tr>
<tr>
<td>H7 The availability of data repository would negatively affects a social scientist’s perceived effort</td>
<td>Supported</td>
<td>−0.176**</td>
</tr>
<tr>
<td>H8 The availability of data repository would positively affects a social scientist’s intention to reuse other scientists’ data</td>
<td>NOT Supported</td>
<td>0.091</td>
</tr>
<tr>
<td>H9 Attitude toward data reuse would positively affects a social scientist’s intention to reuse other scientists’ data</td>
<td>Supported</td>
<td>0.467**</td>
</tr>
</tbody>
</table>

Notes: Unstandardized β, ***p < 0.001, **p < 0.01, *p < 0.05.
The present, novel, study offers a theoretical framework integrating TPB and TAM to explain researchers’ data reuse behavior. The combination of TPB and TAM provides a more comprehensive theoretical framework for studying social scientists’ data reuse, addressing both the social-psychological and practical considerations driving data reuse adoption or rejection by individuals. The results demonstrate that the framework is well supported by the survey data. The combined model using TPB and TAM offers a broader understanding that neither theory could provide independently. TPB alone cannot account for social scientists’ data reuse behaviors, and TAM is useful in explaining how perceived usefulness and effort expectancies influence social scientists’ attitude toward data reuse and their intentions to reuse others’ data. Along with perceived usefulness and effort, perceived concern about data reuse was added to the model, and the relationship between perceived concern and attitude toward data reuse received empirical support from the survey data. This study showed a significant amount of variance in social scientists’ attitudes toward data reuse, which is explained by the perceived usefulness of reused data, the concerns involved in data reuse, and communities’ subjective norms regarding data reuse. Those diverse factors can improve the explanatory power of research into social scientists’ attitude toward data reuse. Furthermore, this study validated the research model based on the survey data with social scientists; the validation of this model points to the potential of its generalizability as a framework for understanding social science data reuse.

7.2. Practical implications

The findings also have several practical implications for stakeholders interested in promoting and facilitating data reuse. First, it is critical to enhance social scientists’ awareness of the potential and benefits of data reuse while decreasing their concerns regarding data reuse. While decisions regarding data reuse are ultimately influenced by individual researchers’ interests and needs, reaching out to the broader communities of researchers, particularly those who have not reused existing data yet, would help researchers understand how data reuse can improve their research and productivity. While research libraries have started to design and provide instructions for data reuse, they need to be more proactive in informing and educating researchers and addressing researchers’ concerns regarding data reuse. Research communities should also be involved in creating workshops or training sessions on data reuse and sharing data reuse experiences with other researchers; in addition to educating researchers on the usefulness of data reuse, education by members of one’s own research community will enforce a subjective community norm of data reuse.

Given the centrality of that subjective norm of data reuse, data reuse is not just driven by personal motivation and interests. Rather, data reuse is also driven by the social norms in disciplines or research communities; those norms are best understood as an informal understanding of the codes of conduct and behaviors that individuals can enact (Lapinski & Rimal, 2005). Nurturing a culture of data reuse would take time, especially if a discipline or community has a tradition of not allowing data reuse. Considering that more and more disciplines are becoming open to data reuse (with rationales as to why reusing existing data is methodologically acceptable) and laws, policies, and regulations have been developed which support it, social scientists may need to revisit the research traditions in their disciplines, including addressing methodological and/or ethical concerns, in order to promote a culture of data reuse. Research support from libraries and repositories will also be essential in creating a subjective norm of data reuse in social science disciplines, as each discipline is still subject to the broader norms of its institutions.

The role of data repositories stood out as encouraging data reuse by its generalizability as a framework for understanding social scientists’ data reuse behaviors. The practices of data reuse can be different in each individual researcher and how researchers interact with what types and formats of data can vary person to person. Follow-up qualitative research would be useful, as it would complement this study, and would offer a more refined understanding of social scientists’ data reuse behaviors.

7.3. Limitations

This study has some methodological limitations. The survey method is limited because it does not capture much contextual information about social scientists’ data reuse behaviors. The practices of data reuse can be different in each individual researcher and how researchers interact with what types and formats of data can vary person to person. Follow-up qualitative research would be useful, as it would complement this study, and would offer a more refined understanding of social scientists’ data reuse behaviors.

For pragmatic reasons explained above, data reuse intentions were used as a proxy for actual data reuse behaviors. Although many research studies support the use of intentions as proxies for actual behaviors (Chen & Chen, 2009; Kuo & Young, 2008; Tsai & Cheng, 2010), there is a possibility that data reuse intentions may not explain social scientists’ actual data reuse behaviors. Because data reuse is a complex and ongoing process in which unexpected challenges may arise (Zimmerman, 2008), expert and novice data reusers’ perceptions may vary. Thus, future studies should measure actual behaviors of data reuse rather than intentions to reuse data, to build on the groundwork laid here and provide a richer understanding of data reuse in the social sciences.

8. Conclusion

This study investigated factors that influence social scientists’ data reuse behaviors using a theoretical model combining TPB and TAM. The findings of this study make an important contribution to the data reuse literature by providing a broad explanation of the relationships among factors that influence data reuse behaviors. This study suggests that subjective norm of data reuse, attitudinal beliefs (e.g., perceived usefulness and concern) and attitude toward data reuse, perceived effort involved in data reuse, and availability of data repository should be taken into account in any effort to facilitate social scientists’ data reuse. Social scientists’ data reuse behaviors can be addressed by lessening the concerns and efforts involved in data reuse, as well as emphasizing the usefulness of data reuse, increasing positive norms of data reuse, and providing reachable data repositories. From the standpoint of librarianship, academic libraries can support social scientists’ data reuse behaviors by providing data search and management services, consulting on copyright and ethical issues in data reuse, and educating researchers about data reuse practices in the social sciences. The findings of this study can encourage scientific data sharing and reuse in social science communities. This research is a step toward enabling data-intensive research in social sciences by suggesting means of supporting and promoting the reuse of shared and accumulated data sets.
Acknowledgements

The authors would like to acknowledge the ProQuest Pivot for allowing the use of its Community of Scientists Scholar Database in recruiting the survey participants, and thank Darra Hofman for reviewing the paper.

Appendix

Appendix A

Measurement items for research constructs.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td>Reusing other researchers’ data improves the quality of my research</td>
<td>(Bock, Zmud, Kim, &amp; Lee, 2005; Kim &amp; Stanton, 2016; Wasko &amp; Faraj, 2000)</td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data enhances the effectiveness of my research</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data reduces the time/cost/effort I spend on my research</td>
<td></td>
</tr>
<tr>
<td>Perceived concern</td>
<td>If I reuse other researchers’ data, I worry that I might misinterpret the data</td>
<td>(Featherman &amp; Pavlou, 2003; Kim &amp; Stanton, 2016; Pavlou, 2003)</td>
</tr>
<tr>
<td></td>
<td>If I reuse other researchers’ data, I worry that I might cause infringement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>If I reuse other researchers’ data, I worry that I might not publish with that data</td>
<td></td>
</tr>
<tr>
<td>Perceived effort</td>
<td>Reusing other researchers’ data requires time and effort to locate data sets</td>
<td>(Davis, 1989; Kim &amp; Stanton, 2016; Thompson, Higgins, &amp; Howell, 1991)</td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data requires time and effort to access (or get permission to use) data sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data requires time and effort to process data sets for a new study</td>
<td></td>
</tr>
<tr>
<td>Attitude toward data reuse</td>
<td>Reusing other researchers’ data is valuable</td>
<td>(Ajzen &amp; Fishbein, 2005)</td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data is desirable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reusing other researchers’ data is pleasant</td>
<td></td>
</tr>
<tr>
<td>Subjective norm of data reuse</td>
<td>In my discipline, it is expected that researchers could reuse other researchers’ data</td>
<td>(Kim &amp; Stanton, 2016; Kostova &amp; Roth, 2002; Son &amp; Benbasat, 2007)</td>
</tr>
<tr>
<td></td>
<td>In my discipline, many of researchers currently reuse data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>In my discipline, reusing other researchers’ data is a common practice</td>
<td></td>
</tr>
<tr>
<td>Availability of data repository</td>
<td>In my discipline, data repositories are available for researchers to share data</td>
<td>(Kim &amp; Stanton, 2016; Taylor &amp; Todd, 1995; Venkatesh, Morris, Davis, &amp; Davis, 2003)</td>
</tr>
<tr>
<td></td>
<td>In my discipline, researchers can easily access data repositories to reuse data</td>
<td></td>
</tr>
<tr>
<td>Intention to reuse data</td>
<td>I am likely to reuse other researchers’ data for my future research</td>
<td>(Ajzen &amp; Fishbein, 2005)</td>
</tr>
<tr>
<td></td>
<td>I intend to reuse other researchers’ data for my future research</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I will try to reuse other researchers’ data for my future research</td>
<td></td>
</tr>
</tbody>
</table>

References


