

# Scientists' Data Reuse Behaviors: A Multilevel Analysis

Youngseek Kim 

*School of Information Science, University of Kentucky, 331 Little Library Building, Lexington, KY 40506, USA.  
E-mail: [youngseek.kim@uky.edu](mailto:youngseek.kim@uky.edu)*

Ayoung Yoon

*School of Informatics and Computing, Department of Library and Information Science, Indiana University-Purdue University, 535 W. Michigan St. IT563, Indianapolis, IN 46202, USA. E-mail: [ayyoon@iupui.edu](mailto:ayyoon@iupui.edu)*

**This study explores the factors that influence the data reuse behaviors of scientists and identifies the generalized patterns that occur in data reuse across various disciplines. This research employed an integrated theoretical framework combining institutional theory and the theory of planned behavior. The combined theoretical framework can apply the institutional theory at the individual level and extend the theory of planned behavior by including relevant contexts. This study utilized a survey method to test the proposed research model and hypotheses. Study participants were recruited from the Community of Science's (CoS) Scholar Database, and a total of 1,528 scientists responded to the survey. A multilevel analysis method was used to analyze the 1,237 qualified responses. This research showed that scientists' data reuse intentions are influenced by both disciplinary level factors (availability of data repositories) and individual level factors (perceived usefulness, perceived concern, and the availability of internal resources). This study has practical implications for promoting data reuse practices. Three main areas that need to be improved are identified: Educating scientists, providing internal supports, and providing external resources and supports such as data repositories.**

## Introduction

Data sharing and reuse produce great benefits for scientists, scholarship, and society, including new research opportunities, validation of existing results, increased efficiency in

Additional Supporting Information may be found in the online version of this article.

**Note:** Both survey data and instrument have been made publicly available via Open ICPSR and can be accessed at <http://doi.org/10.3886/E100404V1>.

Received September 28, 2016; revised January 27, 2017; accepted March 31, 2017

© 2017 ASIS&T • Published online 12 September 2017 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)). DOI: 10.1002/asi.23892

the research cycle, increased research potential by combining new data with reused data, and effective knowledge exchange (Birnholz & Bietz, 2003; Borgman, 2012; National Academy of Science, 2009; Whyte & Pryor, 2011). Despite the wide recognition of these benefits and the push from funders to save time and money in the data re-collection process, there is significant evidence that scientists withhold data for a variety of reasons (Campbell & Bendavid, 2003; Tenopir et al., 2011). Thus, several investments, supports, and regulations have been implemented to encourage scientists to share their data, including: libraries' data services; data management tool development; and data sharing requirements by funders, journals, and other authorities.

The underlying assumption of these discussions and efforts to promote data sharing is that if data are shared, they will be used and reused by other researchers. However, individual scientists have different attitudes toward and experiences with data reuse, because of factors including their different disciplinary cultures, available infrastructure, individual research practices, needs, and awareness. Although some disciplines (e.g., high physics, genomics, and social science) have a longer history of sharing and reusing data through their research infrastructure to support data collection, management, and discovery, in other disciplines, data sharing and reuse tends to occur only through interpersonal exchange (Nelson, 2009; Wallis, Rolando, & Borgman, 2013). Depending on the type of data, some scientists have also debated the ethics of sharing and reusing data and the methodological reasons for not allowing data reuse (Carlson & Anderson, 2007). Tenopir et al. (2015) report that scientists' perceptions and practices of data sharing and reuse have been changing recently; however, data reuse is not yet the norm in every discipline.

In this study, we explored the factors that influence the data reuse behaviors of scientists to identify generalized patterns that occur in those behaviors across various disciplines. Although data reuse studies have been growing in recent years, most of the studies have investigated data reuse

practices within specific disciplines, which limits the potential for understanding the whole picture of scientists' data reuse behaviors across diverse disciplines. Using a theoretical model based on the institutional theory (Scott, 2001) and the Theory of Planned Behavior (Ajzen, 1991), we investigated the factors relating to individual motivations and disciplinary settings that can affect a scientist's intention to reuse data; we also investigated the extent to which these different factors influence those intentions.

## Literature Review

Numerous studies have explored data reuse practices and discussed the barriers to or facilitators of data reuse. These studies address various disciplines, from ecology (Zimmerman, 2008), earthquake science and engineering (Birnholtz & Bietz, 2003; Faniel & Jacobsen, 2010), social science (Niu, 2009; Yoon, 2014b, 2016a), environmental planning (Van House, Butler, & Schiff, 1998), astronomy (Carlson & Anderson, 2007), and archaeology (Faniel, Barrera-Gomez, Kriesberg, & Yakel, 2013), to museum studies (Daniels, 2014). Perhaps because of different disciplinary practices and cultures regarding data reuse, findings are not always consistent (e.g., metadata as the most significant facilitators in neuro anatomical cell researchers' data reuse [Cragin & Shankar, 2006] vs. documentation for earthquake researchers [Faniel & Jacobsen, 2010]), but there some common challenges were also identified across the disciplines.

Many studies that explore data reuse practices suggest that the process of data reuse is not always simple or easy. McCall and Appelbaum (1991) pointed out the amount of time that reusers need to spend absorbing information about data before using it; this is because the reusers are usually unfamiliar with the details of the data. Data reusers in Yoon's (2015) study discussed the learning curve involved in thoroughly understanding new data. Faniel et al.'s (2012) study particularly concerned novice data reusers and their efforts in sense-making in data reuse. Despite that, Niu (2009) argues that data reusers' information needs are the most influential factors in reuse behavior. Niu (2009) found that reusers prioritize their needs the most, and if they think that data are relevant to their research, they will find a way to use the data, even if the process might involve unexpected challenges.

One fundamental challenge in data reuse is that data are embedded in a local context. Because data are the essence of data producers' tacit, theoretical, and technical knowledge, it is often difficult for reusers to understand the data as thoroughly as the researchers, particularly in an interdisciplinary context (Faniel & Zimmerman, 2011). Many studies have reported the complex nature of data (Berg & Goorman, 1999; Cragin & Shankar, 2006; Jirotko et al., 2005). Zimmerman (2008) argues that data reuse creates distance from the original context, which creates problems when reusers try to understand how the data were used. The contextual nature of data also causes problems in transferring context information to data reusers (Birnholtz & Bietz, 2003, p. 341). What contextual information is

important for reusers may depend on the purpose of reuse and the knowledge and experience level of the reusers, but previous studies have suggested some contextual information, including information about the physical, technical, and social environment in which the data were collected, is important to most reusers (Baker & Yarmey, 2009; Chin Jr & Lansing, 2004). Availability of data context information is important for reusers, and reusers are known to search for context information from various sources, such as journals, colleagues, data documentation, data producers, experts, and other reuser communities (Birnholtz & Bietz, 2003; Faniel et al., 2013; McCall & Appelbaum, 1991; Yoon, 2016a).

Often, reusers' distance from the data's original context creates a need to assess data before reusing. Different studies have taken different approaches to understand reusers' assessment of data through the concepts of quality (Cragin & Shankar, 2006; Van House, 2002; Yoon, 2016b; Zimmerman, 2008), reusability (Faniel & Jacobsen, 2010), and trust (Yoon, 2016a). These studies have reported various social, institutional, and individual factors that influence data reuse behaviors and practices.

Finding and obtaining data is still not easy for all data reusers, and Faniel, Kriesberg, and Yakel (2012) found that data accessibility had the strongest relationship with data reusers' satisfaction. Data reusers employ various strategies to find and obtain data because there is no unified system or source for finding data (Rolland & Lee, 2013; Scaffidi, Shaw, & Myers, 2006; Zimmerman, 2008). Thus, the existence and value-adding actions of repositories can also affect data reuse behaviors. Yakel, Faniel, Kriesberg, and Yoon (2013) and Yoon (2014a) found that repositories can enhance the level of reusers' trust in data, which can result in data reuse. Daniels, Faniel, Fear, and Yakel (2012) reported that repositories' value-adding activities—such as correcting errors, creating consistency, responding to the needs of reusers, and reflecting changes in the way that data are collected—support easy access to and reuse of data.

Although these data reuse studies provide valuable insights into data reuse practices and suggest some necessary supports to make the process smoother, many of the findings have limited generalizability because of their methodology (e.g., qualitative approach) or limited study sample (e.g., a specific domain or discipline). This study fills this gap by employing a quantitative approach with a large, multidisciplinary sample. In addition, in contrast to the many previous studies based on researchers' real experiences of data reuse, this study addresses researchers' intentions to reuse data. In this way, we expect that this study will clarify the perceptions of potential data reusers and the factors that influence their behavior.

## Research Model

### *Theoretical Framework*

Because data sharing and reuse practices vary depending on disciplinary context, as well as individual research

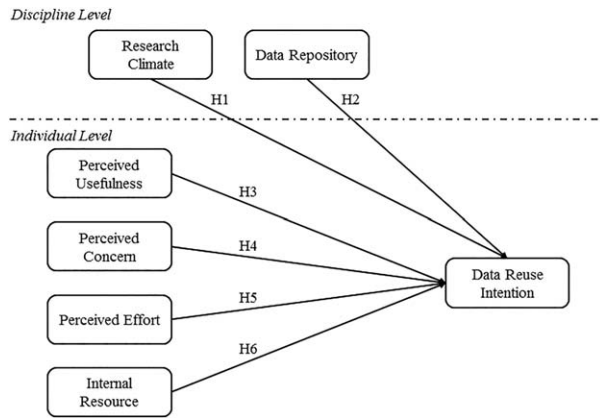


FIG. 1. Multilevel research model for scientists' data reuse intention.

practices (e.g., Borgman, 2007; Pryor, 2009; Tenopir et al., 2011), it is important to investigate both the institutional and individual factors that influence the data reuse behaviors of scientists. This study developed a theoretical model to address both sets of factors while explaining and predicting scientists' data reuse behaviors; this theoretical model combines the institutional theory (Scott, 2001) and the theory of planned behavior (TPB; Ajzen, 1991).

Introduced by Scott (2001), the institutional theory explains the impact of institutional environments on the behavior of individuals (Tolbert, 1985; Tolbert & Zucker, 1983). According to the institutional theory, individuals' beliefs, which become a basis for their actions, are influenced by the institutional environments to which the individuals belong. Thus, their nonrational behaviors are impacted by institutional environments and culture (Powell, 1991; Scott, 2001; Thornton & Ocasio, 2008). Lawrence, Suddaby, and Leca (2011) underscored these diverse institutional influences on an individual's behaviors and argued that individuals interpret which behaviors are legitimately available to them before acting.

However, the institutional theory alone was not sufficient for framing the questions at hand; to investigate both institutional and individual factors in data reuse behavior, our theoretical framework also utilized the theory of planned behavior. Proposed by Ajzen (1991), TPB describes how individuals' beliefs, attitudes, and intentions are related to or influenced by behaviors. The key components of this theory are the individuals' attitudes, subjective norms, and perceived behavioral control that influence the individuals' behaviors through the mediation of behavioral intentions. Each determinant of behavioral intentions is influenced by underlying beliefs, such as behavioral, normative, and control beliefs (Ajzen, 1991; Fishbein & Ajzen, 1975). Combining these two theoretical perspectives provides a useful framework to integrate both institutional and individual perspectives in our research model; institutional theory provides the environmental context of an individual's behavior whereas the TPB describes the underlying motivations behind an individual's behavior.

## Research Model and Hypothesis Development

A research model was developed to explain how disciplinary (or institutional) and individual factors influenced the data reuse behaviors of scientists (Figure 1). At the disciplinary level, the model considered two factors: research climate (i.e., disciplinary research setting) and data repository (i.e., institutional resources). At the individual level, the model considered four factors: perceived usefulness, perceived concern, perceived effort, and the availability of internal, or organizational, resources. The following sections will discuss each research construct and its related hypothesis.

### Disciplinary Level Factors

**Research Climate.** In this study, we defined *research climate* as a disciplinary level factor that influences data reuse by academic scientists. Research climate refers to the scientists' summative evaluation of their academic discipline's research atmosphere in terms of research openness and collaboration. This research focuses on scientists who belong to academic institutions rather than government or corporate scientists who have different research environments in terms of their data practices and policies (Douglass, Allard, Tenopir, Wu, & Frame, 2014). Because the research climate is affected by an academic disciplinary context including sense of community, collaborative relationship, and openness in communication, it can be considered a disciplinary level factor. As data sharing and reuse have increased collaboration and interdisciplinary research (E. S. Lee, McDonald, Anderson, & Tarczy-Hornoch, 2009; Rolland & Lee, 2013; Ross, Lehman, & Gross, 2012), having an open and collaborative disciplinary culture might well influence the data reuse practices in various disciplines. Thus, the research climate within scientific communities would influence a scientist's intention to reuse another scientist's data.

**H1:** *An open, collaborative research climate in a scientific discipline positively influences a scientist's intention to reuse other scientists' data.*

**Availability of Data Repositories.** Having data repositories facilitates data reuse by encouraging scientists to deposit, manage, and make their data available for other researchers. Previous studies demonstrated the roles of data repositories in supporting data reuse: providing access to data, correcting errors, managing provenance, and preserving the data for current and future reuse (e.g., Daniels et al., 2012; Fear & Donaldson, 2012). Repositories also generally enhanced the reusers' trust in the data, which facilitated data reuse (Yakel et al., 2013; Yoon, 2014a). Thus, the availability of data repositories would encourage scientists to reuse other scientists' data.

**H2:** *The availability of data repositories in a scientific discipline positively influences the intention of a scientist within that discipline to reuse other scientists' data.*



### *Individual Level Factors*

*Perceived Usefulness.* The perceived usefulness of data reuse means that reusing other scientists' data provides scientists with benefits in conducting their research, such as an increase in research productivity and effectiveness. Research productivity was one of the major reasons for reusing existing data (Pienta, Alter, & Lyle, 2010). Scientists were aware that reusing existing data could be a cost-effective choice; by saving time and money, data reuse could make it possible to investigate new research problems (Yoon, 2015). Thus, the perceived usefulness of data reuse would encourage scientists to reuse other scientists' data.

**H3:** *The perceived usefulness of data reuse positively influences a scientist's intention to reuse other scientists' data.*

*Perceived Concern.* Perceived concern refers to any potential risks that could be involved in data reuse, including copyright infringement and difficulties publishing results based upon the reused data. Depending on the disciplinary culture and the methodological approaches, data reuse can be perceived as less valuable than research using original data (Goodwin, 2012; Martin, 1995). In some disciplines, it can also be "disconcerting" to allow scientists to freely rely on other scientists' data (Vickers, 2006, p. 15), and researchers reported their difficulties in publishing article that reused existing data (Yoon, 2014b). Thus, the perceived concern involved in data reuse would discourage scientists from reusing other scientists' data.

**H4:** *The perceived concern involved in data reuse negatively influences a scientist's intention to reuse other scientists' data.*

*Perceived Effort.* Perceived effort is the expected effort involved in data reuse. Reusing other scientists' data could involve investing significant amounts of time and effort to retrieve, understand, and process the data in order to make it suitable for another research purpose (Rolland & Lee, 2013; Zimmerman, 2008). Faniel, Kriesberg, and Yakel (2016) also reported that reusers' level of satisfaction was affected by the effort that was necessary to acquire and process the data. Thus, the perceived effort involved in data reuse would discourage scientists from reusing other scientists' data.

**H5:** *The perceived effort involved in data reuse negatively influences a scientist's data reuse intention.*

*Availability of Internal Resources.* "Internal resources" in this research refers to an academic institution's technical and/or human supports for its scientists. Internal resources can better facilitate scientists' data reuse. Studies have reported that reusers experienced difficulties when

reusing data if there was a lack of internal resources (Yoon, 2016a). This research considers the availability of internal resources as individual-level factor rather than discipline-level factor because each scientist has different internal resources depending on their own research environments (e.g., different department/school supports and/or research requirements). Although data reusers can seek external help to overcome these difficulties, the availability of internal resources would facilitate scientists' intentions to reuse other scientists' data. Thus, the availability of internal resources would encourage scientists to reuse other scientists' data.

**H6:** *The availability of internal resources supporting data reuse positively influences a scientist's intention to reuse other scientists' data.*

*Intention to Reuse Data.* It should be noted that, with regards to each factor, this study considered the intention to reuse data as a proxy of actual reuse behavior and measured intention as an outcome variable. Many studies use intention to measure a person's willingness to conduct certain behaviors in various contexts (e.g., Bock, Zmud, Kim, & Lee, 2005; Cho, Chen, & Chung, 2010; Hsu & Lin, 2008; Kuo & Young, 2008a). Studies that used TPB suggested that there was a strong connection between intentions and actual behaviors (Hagger & Chatzisarantis, 2009; Kuo & Young, 2008b; Ramayah, Yeap, & Ignatius, 2013; Tsai & Cheng, 2010). Given these findings in the literature, measuring intentions as an outcome variable of behaviors is methodologically and theoretically sound.

### **Research Method**

This study utilized a survey method to test the proposed hypotheses and to empirically evaluate the research model. A multilevel analysis method was used to investigate the nested nature of scientists within disciplines (i.e., like students within schools) by simultaneously examining both the disciplinary level and the individual level influences on the data reuse behavior of scientists. Because this research utilized the combined theoretical framework integrating institutional theory (discipline level) and theory of planned behavior (individual level), a multilevel analysis is an appropriate analysis method for the nested structure of data including disciplinary and individual level survey items to measure both disciplinary and individual level variances separately (Dansereau, Yammarino, & Markham, 1995; Klein, Dansereau, & Hall, 1994).

### *Population and Sampling*

The study samples included researchers with PhD degrees, such as faculty members and postdoctoral researchers, who were involved in STEM (Science, Technology, Engineering, and Mathematics) disciplines at academic institutions in the U.S. The participants were identified from the

Community of Science's (CoS) Scholar Database (<http://pivot.cos.com>), which included a global research profile directory from only academic institutions, excluding any government or corporate researchers. A total of 15,703 scientists were initially identified for this study using random selection of scientists from the 56 STEM disciplines categorized by CoS—about 280 scientists were selected from each of the 56 STEM disciplines for this study.

### Measurement Scales

The measurement constructs were developed from two theories, and the scales were refined and validated through the process of instrument development (e.g., subject matter experts' review, pre-test, and pilot-test). Most of the items for each construct were developed from previous studies with a slight modification for context, although some items were newly created and validated along with the existing items. All the items were measured using a Likert scale, in which respondents are given a range of 1 to 7 to indicate how strongly they disagree or agree with each statement (item) (see Appendix A for the measurement items for research constructs).

The scores from the multiple measurement items for six independent variables and one dependent variable were averaged to provide an overall score for each of seven variables by each scientist. Then, the two discipline-level independent variables were calculated by aggregating a set of individual scientists' responses in each discipline to each discipline-level variable (Klein & Kozlowski, 2000). The four individual-level independent variables were the same as the average scores from the multiple items for each individual-level independent variable by each scientist. A multilevel regression analysis was then conducted, using these newly developed scores for each variable.

### Data Collection Procedure

An online survey was distributed to the scientists through Qualtrics. The initial email invitation to the survey was sent to 15,703 scientists within academic institutions on October 5, 2015, with a reminder sent on November 10, 2015. The survey closed on November 30, 2015. 1,987 email messages (12.65%) were returned and a total of 13,716 participants (87.35%) received the email invitation to participate in the survey. This research used the National Science Foundation (NSF) STEM discipline codes (2014) for the respondents to indicate their specific academic disciplines based on their current research activities. Of these participants, 1,528 scientists from 94 specific disciplines (as categorized by NSF STEM discipline codes [2014]), completed the survey with less than 5% of missing values (response rate: 11.14%).

Because this research employed a multilevel analysis, any discipline which had less than 10 scientists was excluded from the final data analysis. Two hundred seven scientists from 41 specific disciplines (based on the NSF

TABLE 1. Demographics of survey participants.

Demographic category		Number	Percentage	
Gender	Male	894	72.3	
	Female	324	26.2	
	Missing	19	1.5	
Age	25–34	52	4.2	
	35–44	226	18.3	
	45–54	314	25.4	
	55–64	376	30.4	
	65+	257	20.8	
	Missing	12	1.0	
Ethnic	Asian/Pacific Islander	112	9.1	
	Black/African-American	15	1.2	
	Caucasian	997	80.6	
	Hispanic	39	3.2	
	Native American/Alaska Native	6	0.5	
	Other/Multi-Racial	28	2.3	
	Missing	40	3.2	
Education	PhD/Doctoral Degree	1237	100	
	Status	Tenured	785	63.5
		On Tenure Track	77	6.2
		Not On Tenure Track	254	20.5
		Retired	93	7.5
Missing	28	2.3		
Position	Assistant Professor	90	7.3	
	Associate Professor	268	21.7	
	Full Professor	548	44.3	
	Professor Emeritus	88	7.1	
	Professor of Practice	4	0.3	
	Lecturer/Instructor	24	1.9	
	Post-Doctoral Fellow	41	3.3	
	Researcher	106	8.6	
	Other	68	5.5	
Total		1,237	100	

discipline codes, about 5 scientists per discipline) were excluded because the number of respondents within their disciplines were too small, and the discipline-level factors (i.e., research climate and availability of data repository) could not be measured correctly. Among the 1,321 valid responses from the other 53 specific STEM disciplines based on the NSF discipline codes, any student scientists were excluded because they would not have a solid understanding of the discipline-level factors because of lack of experience in academia. Eighty-four student scientists from the 53 selected disciplines were excluded from the final data analysis. Therefore, a total of 1,237 responses from 53 NSF disciplines were used for the final multilevel data analysis.

### Demographics of the Survey Participants

Most survey participants were male (894, 72.3%) and Caucasian (997, 80.6%), which was not surprising in STEM fields. The survey participants ranged in age from their 20s to their 60s, although the age group of 25 to 34 years old had the least number of participants (52, 4.2%). All participants had a PhD degree in their disciplines as it was one of the criteria for participating in the survey; most of the participants were either tenured researchers (785, 63.5%) or on

TABLE 2. Academic disciplines of the survey respondents.

Discipline	Frequency	Percentage
Biological Sciences	431	34.8
Social Sciences	175	14.1
Engineering	156	12.6
Agricultural Sciences	120	9.7
Earth, Atmospheric, and Ocean Sciences	102	8.2
Health Sciences	86	7.0
Physical Sciences	78	6.3
Psychology	76	6.1
Computer Sciences	13	1.1
Total	1,237	100

TABLE 3. Reliability values ( $N = 1,237$ ).

Variable	Number of items	Cronbach's alpha	Number of cases used
Research Climate	3	.885	1,201
Data Repository	2	.893	1,152
Perceived Usefulness	3	.826	1,031
Perceived Concern	3	.765	1,041
Perceived Effort	3	.790	1,064
Internal Resources	3	.814	1,111
Intention to Reuse Data	3	.968	1,224

the tenure track (77, 6.2%). The summary of the survey participants' demographics is shown in Table 1.

The participants belonged to 53 STEM disciplines according to the NSF discipline codes (2014). Most responses were from 13 subdisciplines of Biological Sciences (431, 34.8%), followed by seven subdisciplines of Social Sciences (175, 14.1%) and nine subdisciplines of Engineering (156, 12.6%). The least responses were from Computer Sciences (13, 1.1%). Table 2 presents the general academic disciplines of the survey respondents.

### Scale Assessment

We assessed the measurement scales by using Cronbach's alpha. A total of seven constructs containing twenty items were tested for reliability of measurement. Cronbach's alpha values of more than .70 are considered to have a strong internal consistency (Nunnally & Bernstein, 1994), as those items provide high explanation power. Cronbach's alpha values for the research constructs employed in this research ranged from 0.765 (Perceived Concern) to 0.968 (Intention to Reuse Data), which are more than the minimum recommended value of 0.70 (Nunnally & Bernstein, 1994). Cronbach's alpha values are presented in Table 3, and show that the research constructs in this research meet the required reliability values.

### Nonresponse Analysis

Nonresponse analysis was conducted to check whether there are any significant differences between participating

TABLE 4. Results from unconditional model.

Fixed effect	Coefficient	Standard error	t-Ratio	P-value
Data sharing behavior ( $\gamma_{00}$ )	2.710	0.050	54.552	<.001
Random effect	Variance component	df	Wald Z	P-value
Intercept ( $u_0$ )	0.068	53	3.984	<.001
Level 1 ( $r$ )	1.247		25.245	<.001

respondents and non-respondents. Babbie (1990) suggested the nonresponse analysis method, which compares early responses and late responses by using the late responses as a proxy for nonresponses. The first 20% of respondents participated in the survey right after the first email was sent, and the last 20% of respondents took the survey right before the survey was closed (i.e., after the reminder was sent).

The ANOVA test shows that there are no significant mean differences between the first and last groups of respondents for both discipline-level and individual-level variables including research climate ( $F = 0.01$ ,  $p = .91$ ), data repository ( $F = 2.29$ ,  $p = .13$ ), perceived usefulness ( $F = 0.11$ ,  $p = .74$ ), perceived concern ( $F = 0.35$ ,  $p = 0.55$ ), perceived effort ( $F = 0.27$ ,  $p = .60$ ), internal resources ( $F = 0.76$ ,  $p = .38$ ), and intention to reuse data ( $F = 3.65$ ,  $p = .06$ ). Therefore, the effects of these nonresponses are marginal, so no weighting method for nonresponse bias was used in this research.

## Results

The multilevel regression analysis was conducted using Hierarchical Linear Modeling (HLM) software. We performed the three-step hierarchical linear modeling procedure suggested by Hofmann (1997), including unconditional model, individual model, and multilevel model. The between- and within-disciplinary level variance in scientists' data reuse intentions was evaluated by using Wald Z based on the unconditional model. The Wald Z results for Intercept ( $u_0$ ) and Level 1 ( $r$ ) are 3.984 ( $p < .001$ ) and 25.245 ( $p < .001$ ) respectively. This means that there is statistically significant amount of variance in disciplinary level factors, and we can continue to analyze the data further. Table 4 shows the unconditional model results.

As the next step of the hierarchical linear modeling procedure, the Level 1 model was evaluated with the individual-level factors only. The Level 1 model only included the four individual-level factors: perceived usefulness, perceived concern, perceived effort, perceived internal resources. Based on the individual-level analysis, perceived usefulness ( $\beta = 0.709$ ,  $p < .001$ ) and internal resources ( $\beta = 0.179$ ,  $p < .001$ ) were found to have positive relationships with scientists' data reuse intentions, and perceived concern ( $\beta = -0.197$ ,  $p < .001$ ) was found to have a

TABLE 5. Random-effect results for scientists' data reuse.

Predictors		Step 1	Step 2	Step 3
		Unconditional model	Individual-level predictors only	Adding group-level predictors
Discipline level predictors	Research climate			-0.135
	Data repository			0.166*
	Residual variance ( $\tau_{00}$ )	0.068	0.023	0.020
Individual level predictors	Perceived usefulness		0.709***	0.710***
	Perceived concern		-0.197***	-0.196***
	Perceived effort		0.015	0.020
	Internal resource		0.179***	0.173***
	Residual variance ( $\sigma^2$ )	1.247	0.589	0.589
Within-group $R^2$		0.528		
Between-group $R^2$			0.130	

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

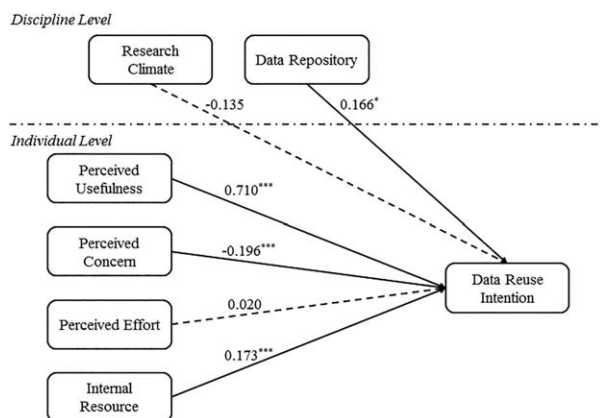


FIG. 2. Hypothesis testing results based on scientists' data reuse intention model.

significant negative relationship with scientists' data reuse intention. However, perceived effort ( $\beta = 0.015, p > .05$ ) was not found to have any significant relationship with scientists' data reuse intentions. The within-discipline variance ( $\sigma^2$ ) changed from 1.247 to 0.589, and the difference presents how much the with-discipline variance accounted for scientists' data reuse intentions based on individual level factors (within-group  $R^2 = 0.528$ ). The three individual level factors including perceived usefulness, perceived concern, and internal resources accounted for 52.8% of the within-discipline variance  $((1.247 - 0.589) / 1.247 = 0.528)$ . Table 5 shows the multilevel analysis results for scientists' data reuse model.

As the final step of the hierarchical linear modeling procedure, the multilevel model was evaluated with both Level 1 (individual level) and Level 2 (discipline level) factors. The multilevel analysis for two discipline-level and four individual-level variables were conducted toward data reuse intention as a dependent variable. At the individual-level, perceived usefulness ( $\beta = 0.710, p < .001$ ) and internal resources ( $\beta = 0.173, p < .001$ ) were found to have positive

relationships with scientists' data reuse intentions, and perceived concern ( $\beta = -0.196, p < .001$ ) was found to have a significant negative relationship with scientists' data reuse intention. However, perceived effort ( $\beta = 0.020, p > .05$ ) was not found to have any significant relationship with scientists' data reuse intentions. At the discipline-level, the availability of data repositories ( $\beta = 0.166, p < .05$ ) was found to have a significant positive relationship with scientists' data reuse intentions; however, research climate was not found to have any significant relationship with scientists' data reuse intentions ( $\beta = -0.135, p > .05$ ). The between-discipline variance ( $\tau_{00}$ ) changed from 0.023 to 0.020, and the difference presents how much the between-discipline variance accounted for scientists' data reuse intentions based on discipline level factors (between-group  $R^2 = 0.130$ ). The discipline level factors including data repository accounted for 13.0% of the between-discipline variance  $((0.023 - 0.020) / 0.023 = 0.130)$ . Figure 2 shows the multilevel analysis results for scientists' data reuse model.

## Discussion

### Discussion of Findings

The results of multilevel analysis show that there are significant between-discipline variances as well as within-discipline variances in the impacts of both individual and disciplinary factors on data reuse intentions. At the individual level, perceived usefulness, perceived concern, and organizational resource were found to have significant relationships with data reuse intention. At the disciplinary level, availability of a data repository was found to have a significant positive relationship with data reuse intention.

Perceived usefulness was found to be the most important factor influencing data reuse intentions, whereas the perceived effort of data reuse is not associated with scientists' intentions to reuse data. This finding aligns with what Niu (2009) reported in her study: for scientists, the usefulness of data is important and they are willing to make efforts to reuse data if they find the data to meet their needs. Thus,



this study confirms that scientists' awareness of the usefulness of data can promote data reuse.

This study did not find any significant relationship between perceived effort and the scientists' intentions to reuse data. Perhaps the needs of the scientist and the usefulness of the data are more important than any effort that might be required to reuse data. However, this finding needs to be explored further. A number of previous studies that examined data reuse practices pointed out that scientists could face a variety of challenges during the process of data reuse and that these challenges were not always obvious to the scientists before they started reusing the data (Faniel, et al. 2012; Scaffidi et al., 2006; Yoon, 2016a; Zimmerman, 2007). Because scientists without data reuse experience might not be fully aware of the effort required for data reuse, further clarification is necessary to confirm how the effort involved in data reuse affects actual data reuse process.

This study shows that scientists' concerns about data reuse (e.g., misinterpretation and infringement) can negatively impact their reuse behaviors. It is well known that scientists have concerns about sharing their data, such as misuse of data by others (e.g., Cragin, Palmer, Carlson, & Witt, 2010; Pryor, 2009), and scientists have similar concerns about reusing others' data. Although previous data reuse studies have reported reusers' concerns about data quality and trustworthiness (e.g., Huang, Stvilia, Jorgensen, & Bass, 2012; Yoon, 2016a, 2016b), not much has been said about reusers' concerns about the risks of misinterpretation and infringement. Although scientists who are experienced in data reuse may have a certain level of confidence (Yoon, 2015), scientists without data reuse experience may have more concerns about misuse of data. As those concerns can be an important impediment to reusing data, it is important to provide proper education and institutional support to address such concerns.

This study suggests that scientists may be more inclined to reuse data when resources and supports for data reuse are easily available at both the disciplinary and individual levels. For instance, resources at scientists' organizations positively influence scientists' intentions to reuse data. Data reuse often requires different levels of human resources for questions (i.e., advisors, data reuser groups, data producers) and technical support to ease the process (i.e., specialized software or programs). Providing these resources and supports can facilitate scientists' comfort in reusing data and help minimize individual efforts in data reuse. Similarly, availability of data repositories also appears to have a significant positive relationship with data reuse intentions. Because searching for and accessing data is the first step of data reuse, repositories can facilitate access to and reuse of data.

### *Practical Implications*

Our findings identified three main areas that are critical to promote data reuse and provided insights into the roles

and contributions of different stakeholders to these areas. The first area is to educate scientists who might be uncertain about or might not be familiar with data reuse. Although data reuse is an individual choice that depends on that individual's research area and appropriate methodologies, knowing about and understanding data reuse could make a difference when designing a research study. Education allows scientists to learn about the potential benefits associated with using existing data and provides answers to any ethical, methodological, and legal concerns that might arise. Many different stakeholders in data reuse can contribute to efforts to educate scientists; research libraries could lead this effort. Research libraries provide many different types of educational services for researchers, including workshops and classes about data management, sharing, and methodologies. These services could be extended to include data reuse and the procedures that are involved in acquiring, processing, and citing existing data.

The second area that was identified as being critical to promoting data reuse practices is internal supports. Various types of internal resources, including experts who can help facilitate the process of data reuse, proper technology, and software and programs that are relevant to data reuse, can influence scientists' intentions to reuse data. Although institutional level supports are based on institutional policy, human resources, and budgets, internal supports are necessary to fulfil the needs of scientists. Libraries can also provide the proper resources that are necessary for data reuse, along with other activities for scholarly supports. A growing number of libraries offer research data services (Briney, Goben, & Zilinski, 2015; Tenopir, Birch, & Allard, 2012) that are relevant to data sharing and management; some of these services relate to data reuse support (e.g., connecting researchers to existing data sources and the list of repositories, and providing data citation standards, etc.; Tenopir et al., 2012). As previous studies found that data reusers seek out help from experts for problem solving during data reuse (Yoon, 2016a); it would also be necessary to provide expert help in data reuse at libraries.

The final area that was identified as being significant in encouraging scientists to reuse data was having external resources and supports, including data repositories. Previous studies recognized the roles of data repositories in data sharing and reuse (e.g., Daniels et al., 2012; Fear & Donaldson, 2012), as well as their impact on reusers (e.g., Yakel et al., 2013; Yoon, 2014a). The findings of this study indicated that the availability of data repositories was an important factor for data reuse. More than 1,500 disciplinary repositories have been registered in re3data (re3data.org, 2016), which shows quantitative growth in the number of data repositories. Although the availability of data through these repositories is a good sign for scientists who seek easy access to existing data, the quality and granularity of the services provided for reusers needs to be further investigated to demonstrate that the available data can meet the reusers' needs and expectations.



## Conclusion

This research showed the big picture of scientists' data reuse across diverse scientific disciplines rather than concentrating on a specific discipline; by expanding the understanding of data reuse across disciplines, this research points towards generalizable factors whose improvement can lead to a broad-based increase in the reuse of data and the attendant gains in efficiency and cost effectiveness. Because scientific data reuse can be influenced by each discipline's institutional environments as well as its scientists' motivations, this research employed an integrated theoretical framework and a multilevel analysis method to better understand scientists' data reuse across disciplines, considering their disciplinary environments and individual motivations together.

Future research is needed to investigate data reuse behaviors in specific disciplines further. This research was of necessity limited in its ability to show any specific data reuse practices in particular disciplines; future research will need examine specific cases of data reuse in particular disciplines further. Also, because this research is quite novel, it included a limited number of research constructs explaining scientists' data reuse intentions; future research is necessary to examine more disciplinary and/or individual level constructs to better understand scientists' data reuse behaviors. In particular, future research should investigate whether scientists' data reuse behaviors are affected by any institutional pushes in their academic disciplines. In addition, future research is needed to examine measuring scientists' actual data reuse behaviors rather than their intentions to reuse data. Research in data reuse can help us to better apprehend scientists' data sharing and reuse behaviors, and the findings of future research can facilitate scientific data sharing and reuse; eventually, this research will advance data-intensive scientific research based on the shared and reused data among scientists.

## Acknowledgments

We would like to acknowledge the CoS Pivot (ProQuest) for allowing us to use their scholar database in recruiting the survey participants. We also would like to acknowledge Ms. Darra Hofman for reviewing this article.

## References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Process*, 52, 179–211.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change* (pp. 173–221). Mahwah, NJ: Lawrence Erlbaum Associates.
- Babbie, E. (1990). *Survey research methods*. Belmont, CA: Wadsworth Publishing.
- Baker, K.S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, 2, 12–27.

- Berg, M., & Goorman, E. (1999). The contextual nature of medical information. *International Journal of Medical Informatics*, 56, 51–60. doi:10.1016/s1386-5056(99)00041-6
- Birnholtz, J.P., & Bietz, M.J. (2003). Data at work: Supporting sharing in science and engineering. Paper presented at the ACM Conference on Supporting Group Work, Sanibel Island, FL.
- Bock, G.W., Zmud, R.W., Kim, Y.G., & Lee, J.N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 29, 87–111.
- Borgman, C.L. (2007). *Scholarship in the digital age: Information, infrastructure, and the internet*. Cambridge: MIT Press.
- Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63, 1059–1078. doi:10.1002/asi.22634
- Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication*, 3, 1–25.
- Campbell, E.G., & Bendavid, E. (2003). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *J Health Care Law Policy*, 6, 241–255.
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12, 635–651.
- Chin, G. Jr, & Lansing, C.S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. Paper presented at the ACM Conference on Computer Supported Cooperative Work, Chicago, Illinois.
- Cho, H.C., Chen, M.H., & Chung, S.Y. (2010). Testing an integrative theoretical model of knowledge-sharing behavior in the context of Wikipedia. *Journal of the American Society for Information Science and Technology*, 61, 1198–1212. doi:10.1002/asi.21316
- 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, 4023–4038. doi:10.1098/rsta.2010.0165
- Cragin, M.H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work*, 15, 185–204.
- Daniels, M., Faniel, I.M., Fear, K., & Yakel, E. (2012). Managing fixity and fluidity in data repositories. Paper presented at the Proceedings of the 2012 iConference.
- Daniels, M. (2014). *Data reuse in museum contexts: Experiences of archaeologists and botanists* (Doctor of Philosophy Dissertation), University of Michigan, Ann Arbor, MI.
- Dansereau, F., Yammarino, F.J., & Markham, S.E. (1995). Leadership: The multiple-level approaches. *The Leadership Quarterly*, 6, 97–109.
- Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance in information technology. *MIS Quarterly*, 13, 319–340.
- Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35, 982–1003.
- Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among full-time government employees. *Journal of the Association for Information Science and Technology*, 65, 251–262. doi: 10.1002/asi.22988
- Faniel, I.M., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A comparative study of data reuse among quantitative social scientists and archaeologists. Paper presented at the iConference, Fort Worth, TX.
- Faniel, I.M., & Jacobsen, T.E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work-the Journal of*

- Collaborative Computing, 19, 355–375. doi:10.1007/s10606-010-9117-8
- Faniel, I.M., Kriesberg, A., & Yakel, E. (2012). Data reuse and sense-making among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49, 1–10.
- Faniel, I.M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67, 1404–1416. doi:10.1002/asi.23480
- Faniel, I.M., & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, 6, 58–69.
- Fear, K., & Donaldson, D.R. (2012). Provenance and credibility in scientific data repositories. *Archival Science*, 12, 319–339.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior*. Reading, MA: Addison-Wesley.
- Goodwin, J. (2012). *The secondary analysis of qualitative data*. Thousand Oaks, CA: Sage.
- Hagger, M.S., & Chatzisarantis, N.L.D. (2009). Integrating the theory of planned behaviour and self-determination theory in health behaviour: A meta-analysis. *British Journal of Health Psychology*, 14, 275–302. doi:10.1348/135910708x373959
- Hofmann, D.A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23, 723–744.
- Hsu, C.-L., & Lin, J.C.-C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, 45, 65–74. doi:10.1016/j.im.2007.11.001
- Huang, H., Stvilia, B., Jorgensen, C., & Bass, H.W. (2012). Prioritization of data quality dimensions and skills requirements in genome annotation work. *Journal of the American Society for Information Science and Technology*, 63, 195–207. doi:10.1002/asi.21652
- Jirotko, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., ... Voss, A. (2005). Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work: CSCW: An International Journal*, 14, 369–398.
- Kim, Y., & Stanton, J.M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 67, 776–799. doi:10.1002/asi.23424
- Klein, K.J., Dansereau, F., & Hall, R.J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, 19, 195–229.
- Klein, K.J., & Kozlowski, S.W.J. (2000). From Micro to Meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, 3, 211–236.
- Kuo, F.Y., & Young, M.L. (2008a). Predicting knowledge sharing practices through intention: A test of competing models. *Computers in Human Behavior*, 24, 2697–2722. doi:10.1016/j.chb.2008.03.015
- Kuo, F.Y., & Young, M.L. (2008b). A study of the intention - Action gap in knowledge sharing practices. *Journal of the American Society for Information Science and Technology*, 59, 1224–1237. doi:10.1002/asi.20816
- Lawrence, T., Suddaby, R., & Leca, B. (2011). Institutional work: Refocusing institutional studies of organization. *Journal of Management Inquiry*, 20, 52–58. doi:10.1177/1056492610387222
- Lee, E.S., McDonald, D.W., Anderson, N., & Tarczy-Hornoch, P. (2009). Incorporating collaborative concepts into informatics in support of translational interdisciplinary biomedical research. *International Journal of Medical Informatics*, 78, 10–21. doi:10.1016/j.ijmedinf.2008.06.011
- Lee, M.-C. (2009). Predicting and explaining the adoption of online trading: An empirical study in Taiwan. *Decision Support Systems*, 47, 133–142.
- Littler, D., & Melanthiou, D. (2006). Consumer perceptions of risk and uncertainty and the implications for behaviour towards innovative retail services: The case of internet banking. *Journal of Retailing and Consumer Services*, 13, 431–443.
- Martin, P.A. (1995). More replication studies needed. *Applied Nursing Research*, 8, 102–103.
- McCall, R.B., & Appelbaum, M.I. (1991). Some issues of conducting secondary analyses. *Developmental Psychology*, 27, 911–917. doi:10.1037/0012-1649.27.6.911
- National Academy of Science. (2009). Ensuring the integrity, accessibility, and stewardship of research data in the digital age. Retrieved from [http://www.nap.edu/catalog.php?record\\_id=12615](http://www.nap.edu/catalog.php?record_id=12615)
- Nelson, B. (2009). Data sharing: Empty archives. *Nature*, 461, 160–163. doi:10.1038/461160a
- Niu, J. (2009). Overcoming inadequate documentation. Paper presented at the American Society for Information Science and Technology, Vancouver, British Columbia, Canada.
- NSF. (2014, February 9). Crosswalk of NSF Discipline Codes to CASPAR Academic Discipline Codes. Retrieved from <https://ncesdata.nsf.gov/nsf/srs/webcasp/data/gradstud.htm>
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.
- Pienta, A.M., Alter, G.C., & Lyle, J.A. (2010). The enduring value of social science research: The use and reuse of primary research data. Paper presented at the organisation, economics and policy of scientific research workshop, Torino, Italy. <http://deepblue.lib.umich.edu/handle/2027.42/78307>
- Powell, W.W. (1991). *Expanding the scope of institutional analysis*. Chicago: University of Chicago Press.
- Pryor, G. (2009). Multi-scale data sharing in the life sciences: Some lessons for policy makers. *International Journal of Digital Curation*, 4, 17–82.
- Ramayah, T., Yeap, J.A.L., & Ignatius, J. (2013). An empirical inquiry on knowledge sharing among academicians in higher learning institutions. *Minerva*, 51, 131–154. doi:10.1007/s11024-013-9229-7
- re3data.org. (2016, April 13, 2016). re3data.org Reaches a Milestone & Begins Offering Badges.
- Rolland, B., & Lee, C.P. (2013). Beyond trust and reliability: Reusing data in collaborative cancer epidemiology research. Paper presented at the Computer Supported Cooperative Work, New York, NY.
- Ross, J.S., Lehman, R., & Gross, C.P. (2012). The importance of clinical trial data sharing toward more open science. *Circulation-Cardiovascular Quality and Outcomes*, 5, 238–240. doi:10.1161/circoutcomes.112.965798
- Scaffidi, C., Shaw, M., & Myers, B. (2006). Games programs play: Obstacles to data reuse. Paper presented at the 2nd Workshop on End User Software Engineering (WEUSE).
- Scott, W.R. (2001). *Institutions and organizations* (2nd ed.). Thousand Oaks, CA: Sage.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *Plos One*, 6, e21101.
- Tenopir, C., Birch, B., & Allard, S. (2012). Academic libraries and research data services. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.3251&rep=rep1&type=pdf>
- Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *Plos One*, 10, e0134826. doi:10.1371/journal.pone.0134826
- Thompson, R.L., Higgins, C.A., & Howell, J.M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly: Management Information Systems*, 15, 125–142.
- Thornton, P.H., & Ocasio, W. (2008). Institutional logics. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin-Andersson (Eds.), *The sage handbook of organizational institutionalism* (Vol. 99–129). Thousand Oaks, CA: Sage.
- Tohidinia, Z., & Mosakhani, M. (2010). Knowledge sharing behaviour and its predictors. *Industrial Management & Data Systems*, 110, 611–631. doi:10.1108/02635571011039052
- Tolbert, P.S. (1985). Institutional environments and resource dependence: Sources of administrative structure in institutions of higher education. *Administrative Science Quarterly*, 30, 1–13.

- Tolbert, P.S., & Zucker, L.G. (1983). Institutional sources of change in the formal structure of organizations: The diffusion of civil service reform, 1880–1935. *Administrative Science Quarterly*, 28, 22–39.
- Tsai, M.T., & Cheng, N.C. (2010). Programmer perceptions of knowledge-sharing behavior under social cognitive theory. *Expert Systems with Applications*, 37, 8479–8485. doi:10.1016/j.eswa.2010.05.029
- Van House, N.A. (2002). Digital libraries and practices of trust: Networked biodiversity information. *Social Epistemology*, 16, 99–114.
- Van House, N.A., Butler, M.H., & Schiff, L.R. (1998). Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 335–343.
- Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 425–478.
- Vickers, A.J. (2006). Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, 7, 15.
- Wallis, J.C., Rolando, E., & Borgman, C.L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *Plos One*, 8, e67332. doi:10.1371/journal.pone.0067332
- Whyte, A., & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *International Journal of Digital Curation*, 6, 199–213.
- Yakel, E., Faniel, I.M., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *International Journal of Digital Curation*, 8, 143–156.
- Yoon, A. (2014a). End users' trust in data repositories: Definition and influences on trust development. *Archival Science*, 14, 17–34.
- Yoon, A. (2014b). "Making a square fit into a circle": Researchers' experiences reusing qualitative data. Paper presented at the American Society for Information Science and Technology, Seattle, WA.
- Yoon, A. (2015). *Data Reuse and Users' Trust Judgments: Toward Trusted Data Curation* (Doctor of Philosophy Dissertation), University of North Carolina at Chapel Hill, Chapel Hill, NC.
- Yoon, A. (2016a). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, 68, 946–956.
- Yoon, A. (2016b). Visible evidence of invisible quality dimensions and the role of data management. Paper presented at the iConference, Philadelphia, PA.
- Zimmerman, A. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7, 5–16.
- Zimmerman, A. (2008). New knowledge from old data - The role of standards in the sharing and reuse of ecological data. *Science Technology & Human Values*, 33, 631–652. doi:10.1177/0162243907306704

## APPENDIX

### Measurement Items for Research Constructs

Construct	Items	Sources
Research Climate	<ul style="list-style-type: none"> <li>• Researchers in my discipline cooperate well.</li> <li>• Researchers in my discipline have a strong feeling of community.</li> <li>• My discipline provides open communication among colleagues.</li> </ul>	Bock et al. (2005), Tohidinia and Mosakhani (2010)
Availability of Data Repositories	<ul style="list-style-type: none"> <li>• In my discipline, data repositories are available for researchers to share data.</li> <li>• In my discipline, researchers can easily access data repositories to reuse data.</li> </ul>	Kim and Stanton (2016), Venkatesh, Morris, Davis, and Davis (2003)
Perceived Usefulness	<ul style="list-style-type: none"> <li>• Reusing other researchers' data improves the quality of my research.</li> <li>• Reusing other researchers' data enhances the effectiveness of my research.</li> <li>• Reusing other researchers' data reduces the time/cost/effort I spend on my research.</li> </ul>	Davis (1989), Davis, Bagozzi, and Warshaw (1989)
Perceived Concern	<ul style="list-style-type: none"> <li>• If I reuse other researchers' data, I worry that I might misinterpret the data.</li> <li>• If I reuse other researchers' data, I worry that I might cause infringement.</li> <li>• If I reuse other researchers' data, I worry that I might not publish with that data.</li> </ul>	M.-C. Lee (2009), Littler and Melanthiou (2006)
Perceived Effort	<ul style="list-style-type: none"> <li>• Reusing other researchers' data requires time and effort to locate data sets.</li> <li>• Reusing other researchers' data requires time and effort to access (or get permission to use) data sets.</li> <li>• Reusing other researchers' data requires time and effort to process data sets for a new study.</li> </ul>	Davis et al. (1989), Thompson, Higgins, and Howell (1991)
Availability of Internal Resources	<ul style="list-style-type: none"> <li>• In my organization (e.g., university), resources that promote data reuse are available to me.</li> <li>• Information technologies are available to support my data reuse.</li> <li>• A specialized person is available to assist with my data reuse.</li> </ul>	Thompson et al. (1991), Venkatesh et al. (2003)
Data Reuse Intention	<ul style="list-style-type: none"> <li>• I am likely to reuse other researchers' data for my future research.</li> <li>• I intend to reuse other researchers' data for my future research.</li> <li>• I will try to reuse other researchers' data for my future research.</li> </ul>	Ajzen and Fishbein (2005), Tohidinia and Mosakhani (2010)