View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Kentucky

University of Kentucky
UKnowledge

Information Science Faculty Publications

Information Science

4-2016

# Institutional and Individual Factors Affecting Scientists' Data-Sharing Behaviors: A Multilevel Analysis

Youngseek Kim
*University of Kentucky*, youngseek.kim@uky.edu

Jeffrey M. Stanton
*Syracuse University*

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

Follow this and additional works at: https://uknowledge.uky.edu/slis_facpub

Part of the Scholarly Communication Commons

## Repository Citation

Kim, Youngseek and Stanton, Jeffrey M., "Institutional and Individual Factors Affecting Scientists' Data-Sharing Behaviors: A Multilevel Analysis" (2016). *Information Science Faculty Publications*. 16.
https://uknowledge.uky.edu/slis_facpub/16

This Article is brought to you for free and open access by the Information Science at UKnowledge. It has been accepted for inclusion in Information Science Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

**Institutional and Individual Factors Affecting Scientists' Data-Sharing Behaviors: A Multilevel Analysis**

**Notes/Citation Information**
Published in the *Journal of the Association for Information Science and Technology*, v. 67, issue 4, p. 776–799.

© 2015 ASIS&T

This is the accepted version of the following article: 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, which has been published in final form at http://dx.doi.org/10.1002/asi.23424. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**Digital Object Identifier (DOI)**
http://dx.doi.org/10.1002/asi.23424

This article is available at UKnowledge: https://uknowledge.uky.edu/slis_facpub/16

**This is the accepted version of the following article: 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, which has been published in final form at http://dx.doi.org/10.1002/asi.23424. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.**

**Institutional and Individual Factors Affecting Scientists' Data Sharing Behaviors:**

**A Multilevel Analysis**

Youngseek Kim, Ph.D.*

University of Kentucky

School of Library and Information Science

331 Little Library Building

Email: youngseek.kim@uky.edu

Phone: (859) 218-2295

Fax: (859) 257-4205

Jeffrey M. Stanton, Ph.D.

Syracuse University

School of Information Studies

206 Hinds Hall

Email: jmstanto@syr.edu

Phone: (315) 443-2879

Fax: (315) 443-5806

## ABSTRACT

The objective of this research was to investigate the institutional and individual factors that influence scientists' data sharing behaviors across different scientific disciplines. Two theoretical perspectives, institutional theory and theory of planned behavior, were employed in developing a research model that showed the complementary nature of the institutional and individual factors influencing scientists' data

* Corresponding Author

sharing behaviors. This research used a survey method to examine to what extent those institutional and individual factors influence scientists' data sharing behaviors in a range of scientific disciplines. A national survey (with 1,317 scientists in 43 disciplines) showed that regulative pressure by journals; normative pressure at a discipline level; and perceived career benefit and scholarly altruism at an individual level had significant positive relationships with data sharing behaviors; and that perceived effort had a significant negative relationship. Regulative pressure by funding agencies and the availability of data repositories at a discipline level and perceived career risk at an individual level were not found to have any significant relationships with data sharing behaviors.

Keywords

Data sharing, data reuse, eScience, cyberinfrastructure, institutional theory, theory of planned behavior, multilevel analysis

**INTRODUCTION**

Data sharing is a critical issue in modern scientific research with the emergence of e-Science or cyberinfrastructure. e-Science revolutionized the process of scientific discovery by enabling data-centric science or scientists sharing their data through technological development and collaborative effort (Hey & Trefethen, 2008). In addition, primary data collected by individual scientists becomes an important "information currency" along with research analyses and finding in the traditional publications (Davis & Vickery, 2007). As the primary data becomes important in terms of scientific research and scholarly communication, data sharing is now essential in most modern research activities.

In the last few decades, the science and engineering communities made continuous endeavors to promote scientists' data sharing in order to improve scholarly communication and eventually realize the vision of data-centric scientific research. National funding agencies, in order to leverage their investments, began to require their grant awardees to eventually make primary data available to others (National Science Foundation, 2010). Researchers gradually agreed that primary data generated by public funding should be shared with others (Arzberger et al., 2004). Additionally, many scientific journals' data sharing policies began to mandate data sharing for the published articles. Such mandates were implemented throughout several scientific communities (Faniel & Zimmerman, 2011).

Despite continuous efforts by funding agencies and science institutions, data sharing is still not well-deployed throughout science and engineering disciplines. Although data sharing benefits scientists and improves scientific research development, scholars observed that data sharing is not a common practice (Piwowar & Chapman, 2010). In some disciplines, such as genetics and molecular biology, scientists continue to have prolific positive outcomes through data sharing. Still, many other disciplines do not fully deploy the idea of data sharing for their scientists and engineers. Sometimes, even fields that have good support for data sharing still struggle with the actual data sharing by individual scientists.

There are several barriers that prevent scientists from sharing data. According to the traditional norms of science, scientists are supposed to share their scientific findings and related information under the ideals of communalism (Merton, 1973). However, disciplinary traditions, institutional barriers, lack of technological infrastructure, intellectual property concerns, and individual perceptions prevent scientists from sharing their data with others. Prior efforts focused on the development of data repositories and relevant technical tools to facilitate scientists' data sharing. However, diverse external issues, including the policies developed by funding agencies and journals continue to influence scientists' data sharing (Borgman, 2010). Related to these institutional issues, individual scientists' perception toward data sharing significantly influences their data sharing behaviors (Tenopir et al., 2011).

Compared to the importance of data sharing in scientific research, prior studies do not fully address the complex nature of data sharing. Scholars from a diverse range of disciplines studied scientists' data sharing, in order to understand both the prevalence of sharing or withholding of data, and factors that influence data sharing or withholding. Although scientists' data sharing practices are embedded in a higher level context (i.e., scientific discipline or institution), prior studies focused on the technical and the individual aspects of data sharing, rather than combining them within their institutional contexts. The institutional or disciplinary context is critical for understanding scientists' data sharing. Each scientific discipline has its own institutional context(s), influencing its scientists' data sharing behaviors, along with individual and technological aspects of data sharing.

The main objective of this research is to investigate the factors influencing scientists' data sharing behaviors in diverse scientific communities. This research considers both individual and contextual factors in influencing scientists' decisions to share their data with others. More specifically, this research considers the combination of institutional and individual factors that influences scientists' decisions on data sharing behaviors. By taking an integrated perspective both at the disciplinary and individual levels, this research demonstrates the dynamics of institutional and individual influences affecting scientists' data sharing behaviors.

## LITERATURE REVIEW

Data sharing would be desirable scientific behavior under the norms of communalism and disinterestedness proposed by Merton (1973); however, other scholars argued that scientists do not behave entirely according to Merton's scientific norms, but rather also seek their own interests based on solitariness and interestedness (Mitroff, 1974; Mulkay, 1976). A number of later studies also showed the gap between Merton's norms of scientists and their actual behaviors (Cronin, 1984; Kellogg, 2006; Ziman, 2000). Unlike traditional publication methods, data sharing does not have standard or formal mechanisms of citation, and thus cannot provide appropriate rewards for the scientists who collected the data (Borgman, 2010). Therefore, data sharing has not yet been established as major scholarly communication methods throughout different scientific communities (Borgman, 2007; Tenopir et al., 2012).

Although sharing data among scientists means that more scientists can benefit from the data, there is ample evidence that scientists nonetheless withhold their data rather than sharing it in popular science journals (Campbell & Bendavid, 2003; Cohen, 1995; Piwowar, 2011). Previous literature on scientists' data sharing and withholding has paid considerable attention to (1) the prevalence of data sharing and withholding, (2) the motivations behind and barriers to data sharing and withholding, and (3) the benefits and consequences of data withholding (Campbell et al., 2002; Campbell, Louis, & Blumenthal, 1998; Campbell, Weissman, Causino, & Blumenthal, 2000; Kim & Stanton, 2012; Louis, Jones, & Campbell, 2002).

Especially, prior studies research on diverse factors influencing scientists' data sharing and withholding, and those factors can be categorized into three groups: (1) Institutional factors including funding agency's policy (McCullough, McGeary, & Harrison, 2008; Piwowar & Chapman, 2008), journal requirements (McCain, 1995; Piwowar & Chapman, 2008), and contract with industry sponsors (Louis et al., 2002), (2) IT resource factors including metadata (Bietz, Baumer, & Lee, 2010; Hey & Trefethen, 2004; Karasti, Baker, & Millerand, 2010) and data repositories (Choudhury, 2008; Witt, 2008), and (3) individual factors including personal characteristics (Campbell & Bendavid, 2003; Campbell et al., 2002), perceived benefit (Kim, 2007; Kankanhalli, Tan, & Wei, 2005; Kling & Spector, 2003), perceived effort (Campbell et al., 2002; Louis et al., 2002; Tenopir et al., 2011), perceived risk (Reidpath & Allotey, 2001; Savage & Vickers, 2009; Stanley & Stanley, 1988). In addition, other organizational and environmental factors have been studied as important factors influencing scientists' data sharing and withholding (Tenopir et al., 2011; Vogeli et al., 2006). These prior studies informed this research to choose a relevant theoretical framework and develop a concrete research model by addressing institutional, IT resource, and individual factors together.

Previous studies provided valuable insights; however, they were limited in terms of main focus, research methods, theoretical frameworks used, what research constructs were employed, and what disciplines were studied. First, previous studies focused mainly on individual motivational factors and technical factors rather than institutional and environmental factors. Next, the majority of previous studies hardly employed any theoretical model to explain scientists' data sharing behaviors. Additionally, previous studies identified few research constructs regarding the factors influencing scientists' data sharing. Finally, previous studies did not cover diverse science and engineering disciplines in regards to scientists' data sharing behaviors. By understanding the limitations of previous studies, this research discussed possible theoretical framework and research method that can triangulate scientists' data sharing behaviors.

## THEORETICAL FRAMEWORK

Drawing upon institutional theory and the theory of planned behavior, this research proposed a research model to investigate how both institutional and individual drivers influence scientists' data sharing behaviors. Scientists' data sharing behaviors can be understood through the lens of institutions' seeking organizational legitimacy and individual motivation. Institutional theory (Scott, 2001) provides significant insights regarding the importance of institutional environments including institutional rules, norms, and culture on individuals' actions (behaviors) (Tolbert, 1985; Tolbert & Zucker, 1983). In contrast, the theory of planned behavior provides insights regarding how individuals' attitudes, subjective norms, and perceived behavioral control influences individuals' behaviors mediated by intention (Ajzen, 1991).

The research model builds on insights from Scott's (2001) neo-institutional theory. According to Scott (2001), institutions shape individuals' beliefs and their non-rational behaviors. Individuals are embedded in institutional environments; these provide individuals with a basis for actions and shape individuals' priorities and activities (Powell, 1991; Thornton & Ocasio, 2008). Individual actors consider diverse institutional influences in order to interpret what actions are legitimately available to them and make their decisions (Lawrence, Suddaby, & Leca, 2011).

Neo-institutional theory posits three kinds of institutional pressures influencing behaviors: regulative, normative, and cultural-cognitive. These institutional pressures provide guidelines and constrain actions (Scott, 2001, p. 50). Regulative pressure arises from the rules that an authoritative organization or actor sets for desirable behaviors of other organizations or its organizational members. Regulative pressure provides organizations or individuals with coercive constraints, and legally sanctions those who do not comply. Normative pressure refers to social obligation caused by collective expectations in a community.

Normative pressure sets shared norms for the appropriateness of individuals' or organizations' behaviors. Training, education, and association teach individuals shared norms, and individuals are governed morally by these collective expectations. Lastly, cultural-cognitive pressure refers to the shared understanding of the world that is taken for granted. The cultural-cognitive institution is deeply embedded in communities and is supported culturally. Organizations or individuals observe others' activities and simply imitate their behaviors.

The theory of reasoned action and its successor, the theory of planned behavior are well-established social psychology theories that describe how salient beliefs influence behavioral intentions and subsequent behavior (Ajzen, 1991; Fishbein & Ajzen, 1975). Theory of planned behavior explains an individual's behavior based on his or her behavioral intentions. These intentions are in turn influenced by his/her attitude toward a behavior, perception of the subjective norms regarding that behavior, and perceived behavioral control. Behavioral intention refers to a person's aim to perform a particular behavior (Ajzen, 1991). An attitude is a cognitive and emotional evaluation of an object or behavior (Ajzen, 1991). A subjective norm is a person's belief that people who are important to him or her expect that he or she should or should not perform a particular behavior (Ajzen, 1991). Perceived behavioral control is an individual's perceptions of his or her ability to perform a given behavior easily (Ajzen, 1991). Each of the determinants of behavioral intention is in turn influenced by underlying belief structures such as behavioral, normative, and control beliefs (Ajzen, 1991; Fishbein & Ajzen, 1975).

Based on institutional theory and theory of planned behavior, a research model was developed to explain and predict scientists' data sharing behaviors. Drawing on theories and previous literature, this research identifies two groups of factors – institutional predictors and individual predictors, respectively – that influence scientists' data sharing behaviors. The combination of two theoretical perspectives provides an opportunity to examine scientists' data sharing behaviors from both institutional and individual perspectives. Institutional theory explains the context within which individual scientists are acting; whereas the theory of planned behavior explains the underlying motivations behind scientists' data sharing behaviors in an institutional context. The Figure 1 below shows the research model for scientists' data sharing behaviors.

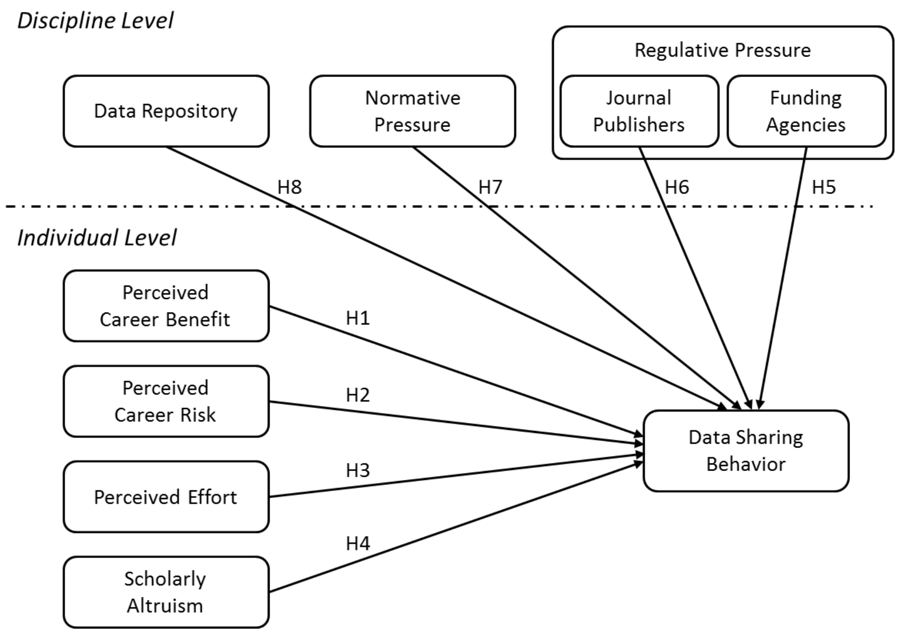



Figure 1. Research Model for Scientists' Data Sharing Behaviors

*Individual Level*

The three behavioral beliefs toward data sharing including perceived career benefit, perceived career risk, and perceived effort would influence scientists' data sharing behaviors. Based on prior literature, we found that these three behavioral beliefs are the main individual level perceptions that either positively or negatively influence scientists' data sharing behaviors.

Perceived Career Benefit

Scientists' perceptions of the career benefit of data sharing would positively influence their data sharing behaviors. Perceived career benefit means the degree to which a scientist believes that sharing data could provide rewards such as recognition and reputation through acknowledgements, citations, and sometimes authorships. Prior studies reported that scientists' perceptions of rewards (i.e., acknowledgements, citations, and authorship) for data sharing enhanced their data sharing behaviors (Kankanhalli et al., 2005;

Kling & Spector, 2003); however, if they perceive low or no reward, they are unlikely to share their data with others (Sterling & Weinkam, 1990). Thus, the perceived career benefit of data sharing would encourage scientists to share their data with other scientists.

H1: The perceived career benefit of data sharing positively influences scientist's data sharing behavior.

Perceived Career Risk

Perceived career risk is defined as a scientist's belief about the potential negative outcomes from data sharing that might affect his or her career adversely. The perception of data sharing as risky is an important barrier for scientists who are considering whether to make their data available to other scientists. In the context of scientists' data sharing, prior studies identified diverse components of perceived (career) risk including losing publication opportunities (Reidpath & Allotey, 2001; Savage & Vickers, 2009; Stanley & Stanley, 1988), protecting one's career (Campbell et al., 2002; Louis et al., 2002), and misuse of data (Borgman, 2007; Cragin, Palmer, Carlson, & Witt, 2010; Pryor, 2009). Therefore, if scientists believe that data sharing has possible negative outcomes for their careers, they are less likely to share their data with others.

H2: The perceived career risk involved in data sharing negatively influences scientist's data sharing behavior.

Perceived Effort

Perceived effort refers to the degree to which a scientist believes that sharing data would require work (energy) and time. In the context of knowledge sharing, Thorn and Connolly (1987) found that individuals were less likely to share their knowledge the more time and effort it took to share it. In regards to scientists' data sharing, prior studies also pointed out time and effort required to share their data impeded scientists' data sharing (Campbell et al., 2002; Stanley & Stanley, 1988; Tenopir et al., 2011). Therefore, if scientists believe that data sharing requires their effort, they are less likely to share their data with others.

H3: The perceived effort required to share data negatively influences scientist's data sharing behavior.

Scholarly Altruism

Scientists' scholarly altruism would increase their data sharing behaviors. Scholarly altruism refers to the degree to which a scientist is willing to work to increase others' welfare without expecting any benefits in return (Hsu & Lin, 2008). There are few prior studies focusing on the link between (scholarly) altruism and scientists' data sharing. A couple of studies found that altruism is an important factor influencing faculty members' contribution to institutional data repositories (Foster & Gibbons, 2005; Kim, 2007). Those faculty members who contribute their data to institutional repositories have greater altruism to make their data available to the public (Cronin, 2005; Foster & Gibbons, 2005; Kim, 2007). Therefore, if scientists have more altruistic motivations, they are more likely to share their data with others.

H4: Scientist's scholarly altruism positively influences his/her data sharing behavior.

Data Sharing Behavior

This research considers actual data sharing behavior as an outcome variable. In the context of scientists' data sharing, data sharing behavior can be defined as the extent to which scientists provide other scientists with their research data and information related to their published articles by depositing them into data repositories and providing them upon request. In this research, data sharing behaviors can be determined by both individual predictors (i.e., perceived career benefit, perceived career risk, perceived effort, and scholarly altruism) and institutional predictors (i.e., regulative pressures by funding agencies and journal publishers, normative pressure, and the availability of data repositories).

*Institutional Level*

Regulative Pressures by Funding Agencies and Journals

Governmental funding agencies and journal publishers require scientists to share data in order to receive funding or publish articles in their journals. Scientific funding agencies create data management and sharing policies requiring grantees to share raw data with others. Funding agencies can increase regulative pressures on scientists by controlling the funding resources available to them. As such, scientists are subject to influence from scientific funding agencies such as NSF and NIH. Because the resources controlled by these funders are so critical to scientists' career progress, they need to comply to secure their own professional survival (Pfeffer & Salancik, 1978).

Similarly, many science and engineering journals in some disciplines require their authors to share original data in various ways, such as submitting data to data repositories, and/or providing data upon request. Since journal publishers control access to the publication of research articles, they are one of the dominant sources of coercion for scientists. Scientists who feel more regulative pressures from journals will be more likely to share their data with others. Prior studies found that the compliance with regulative pressures influence individuals' intention and their actual behaviors directly (Liu, Ke, Wei, Gu, & Chen, 2010; Teo, Wei, & Benbasat, 2003). Therefore, this research assumes that the regulative pressures by funding agencies and journal publishers would directly influence scientists' data sharing behaviors.

H5: The regulative pressure by funding agencies positively influences scientist's data sharing behavior.

H6: The regulative pressure by journal publishers positively influences scientist's data sharing behavior.

Normative Pressure

In the context of scientists' data sharing behaviors, normative pressure would lead scientists who are in the same community to follow the socially adopted norms of their communities. Normative pressures constrain scientists' data sharing behaviors through a system of values, norms, expectations, and roles (DiMaggio & Powell, 1991; Scott, 2001). Ceci (1988) found that scientists in the physical and social sciences endorse the data sharing principle, since it is a desirable norm in scientific communities. Scientists' perceptions of normative pressure originate from their research communities, which share similar values, norms, and expectations. Scientists conform to norms in order to maintain their legitimacy by reassuring constituents in their fields (John, Cannon, & Pouder, 2001; Zsidisin, Melnyk, & Ragatz, 2005). The institutional norm as the forms of professionalism and expectation from peer-scientists in a scientific community would positively influence scientists' data sharing behaviors.

H7: The normative pressure in a scientific discipline positively influences scientist's data sharing behavior.

Data Repository

The institutional resources that are already known as resource-facilitating conditions in prior studies would be important institutional level factors influencing scientists' data sharing. Prior studies found that resource-facilitating conditions reduce the perceived efforts as individual's attitudinal belief (Phang et al., 2006). Resource-facilitating conditions have been studied in prior knowledge sharing studies, and those studies revealed that the resource-facilitating conditions play an important role in predicting people's attitude toward knowledge sharing, intentions to share knowledge (Ryu, Ho, & Han, 2003; So & Bolloju, 2005). Therefore, scientists' resource-facilitating conditions including data repositories would enhance scientists' data sharing behaviors.

H8: The availability of data repositories in a discipline positively influences scientist's data sharing behavior.

The current research focuses on how institutional and individual factors influence scientists' data sharing behaviors across scientific disciplines. The research model and hypotheses developed at this stage were empirically validated by using survey data collected from scientists in diverse science and engineering disciplines. The survey research helps in investigating data sharing factors at individual and institutional levels.

## RESEARCH METHOD

This research employed a survey method to examine the constructs and hypothesized relationships of the scientists' data sharing model. By conducting the survey in a range of science and engineering disciplines, this research validated the scientists' data sharing model by investigating both institutional and individual influences of scientists' data sharing behaviors. In addition, consistent with the multilevel theoretical framework combining institutional theory (discipline level) and theory of planned behavior (individual level), a multilevel analysis was employed for this research, since the estimation of variances in different levels is theoretically relevant (Dansereau, Yammarino, & Markham, 1995; Klein, Dansereau, & Hall, 1994).

Population and Sampling

The target population of this research included faculty members and post-doctoral researchers in U.S. academic institutions who belong to STEM (Science, Technology, Engineering, and Mathematics) disciplines. They are expected to have their own data collected and to have ownership of those data. The sampling frame of this research was identified from the scholar list in the Community of Science's (CoS) Scholar Database (http://pivot.cos.com). This database provides a global researcher profile directory, though mainly from universities and colleges rather than from industry. The CoS scholar database provides the means to directly access the population of this research. Based on the list of scholars who were registered in U.S. academic institutions, scientists were randomly selected from STEM disciplines categorized in the CoS database.

This research originally planned to collect a sample size of at least 40 disciplines, with a minimum of 20 scientists per discipline according to the sample size recommendations of prior studies (Goldstein, 2011; Hox, 2002; Raudenbush & Bryk, 2002; Scherbaum & Ferreter, 2009). Since this research measures group-level variables based on individual data in each group, at least 20 scientists (observations) are needed in each discipline (Hox, 2002; Scherbaum & Ferreter, 2009); however, we decided to include any discipline which has more than 15 scientists for the final data analysis in order to increase the statistical power to detect the discipline-level predictors. Since this research has four Level-2 predictors (i.e., regulative pressures by funding agencies and journals, normative pressure, availability of data repositories), it is necessary to have at least 40 disciplines to detect Level-2 effects (Goldstein, 2011;

Raudenbush & Bryk, 2002). Scholars argued that a sufficient number of groups are required to estimate the level 2 parameters properly (Goldstein 2011; Raudenbush et al. 2002), and it is more important to increase the number of groups included in multilevel analysis as opposed to the number of members in each group (Zhang et al. 2009). By lowering the minimum number of scientists in each discipline from 20 to 15, this research can include more disciplines for its multilevel analysis, and eventually this would increase statistical power and make Level-2 estimates stable.

Measurement of Constructs

The theoretical framework was translated into measurements of constructs. The measurement scales were refined and validated through the instrument development procedure including subject matter experts' review, pre-test, and pilot-test. Most of the survey items were adapted from previous studies, and they were modified for the context of scientists' data sharing through the scale development procedure. Some of the survey items were newly created and validated with the existing measurement items. At the beginning of the survey, we provided a brief definition of data sharing saying that "In this survey, Data Sharing means providing the raw data of your published articles to other researchers outside your research group(s) by making it accessible through data repositories/ public web spaces/ supplementary materials or by sending the data via personal communication methods upon request." With regards to the measurement of scientist's data sharing behavior, new items were developed to capture diverse forms of data sharing behaviors by considering the number of times they share their data with others. In this study, a minimum of three items for each construct were used to measure each construct (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Rakov & Marcoulides, 2000). All the variables were measured using Likert scales (1 – 7), ranging from "Strongly Disagree" to "Strongly Agree" for scientists' perceptions and disciplinary factors regarding their data sharing; or "Never" to "Always" for their data sharing behaviors. Respondents were asked to mark the response that best described their level of agreement in the statements.

Since this research employed a multilevel model, institutional level constructs needed to be measured properly in order to conduct a multilevel analysis. Regulative pressures, normative pressure, and institutional resources in a discipline can be considered as "shared (institutional) properties" because they are usually originated from experience, perceptions, and values (Klein & Kozlowski, 2000). These shared (institutional) property constructs were measured by individual scientists' subjective rating for the items of those constructs. Through these subjective measurements, this research examined the extent to which those shared property constructs are shared by individual scientists in a same discipline (Klein & Kozlowski, 2000). The measurement items for each construct and its sources are indicated in the Appendix A.

Data Collection Procedure and Result

Since this research involved human subjects, Institutional Review Board (IRB) approval was granted prior to data collection. This research was approved by the IRB at authors' research institution. The IRB allowed us to perform the national survey with the sampling frame based on the list of scientists from the CoS scholar database. A formal request was made to receive permission from the CoS Pivot (ProQuest) in conducting a random sampling from its scholar database, and CoS Pivot allowed us to perform the random sampling using their scholar database for only the purpose of this research.

The survey questionnaire was created and distributed to individual scientists by using SurveyGizmo. The online survey questionnaire consists of research introduction and purpose, specific questions to measure the constructs, and respondents' demographic information. This online survey presented an online consent form at the beginning of the survey, so the participants can proceed to this survey by agreeing to the survey requirements by IRB. Once a participant has submitted the survey, the survey data was recorded in the online survey (SurveyGizmo) server and used for the future data analyses. Two incentives were offered for survey participants who submitted their responses and provided their email addresses: (1) a raffle to win one of ten $50 gift cards and (2) the final report of this survey.

The final field survey instrument was distributed to the 16,165 potential survey participants in 56 STEM disciplines. From November 19, 2012 to February 15, 2013, a total of 2,674 participants submitted their partial and full responses. Out of 2,674 responses, there were 2,470 valid responses used for the initial data analysis (15.28% of response rate). From the original 2,674, 204 responses were removed because those responses were missing more than 20% of answers and/or the answers regarding participants' data sharing behaviors. Reports of data sharing behaviors were critical for all aspects of the data analysis. Out of 2,470 initial responses, this research excluded (1) scientists who were from non-academic institutions since their data-sharing decisions may be made by their organizations (298, 12.06%), (2) student scientists, since they often do not have any authority to share their research data and may not have a clear understanding about institutional pressures (e.g., funding agencies' requirement), (247, 10.00%), and (3) the scientists who did not produce any data related to their publications in the last two years, since their responses on both discipline-level and individual-level independent variables cannot lead to its dependent variable (i.e., data sharing behavior) (155, 6.28%). In terms of the number of scientists in each discipline, this research excluded any disciplines that had fewer than 15 qualified scientists (304, 12.31%) or that were categorized as "others" (e.g., bioscience-other) (149, 6.03%). This resulted in 1,317 usable responses for the final data analysis for hypothesis testing, and out of 2,470 initial usable responses, 1,153 responses were excluded.

## Demographics of the Respondents

The descriptive statistics of demographics include gender, age, ethnicity, education, position, status, sector, and discipline. Of the selected sample of 1,317 scientists, there were 936 male participants (71.07%) and 348 female participants (26.42%), while 33 participants (2.51%) did not indicate their gender. In terms of age, the survey participants are well distributed in each age group: 25-34 (139, 10.55%), 35-44 (332, 25.21%), 45-54 (334, 25.36%), 55-64 (328, 24.91%), 65+ (174, 13.21%), and 10 (0.76%) missing values. With regards to the distribution of ethnicity, the number of Asian was 167 (12.68%), African-American was 14 (1.06%), Caucasian was 1,046 (79.42%), Hispanic was 32 (2.43%), Native American was 1 (0.08%), Other/Multi-Racial was 27 (2.05%), and 30 participants (2.28%) did not indicate ethnicity. In terms of position, most of the survey participants were professors. They were listed as full professor (544, 41.31%), associate professor (305, 23.16%), assistant professor (197, 14.96%), professor emeritus (53, 4.02%), professor of practice (6, 0.46%), and lecturer (8, 0.61%). There were also these distinctions in respondents: post-doctoral fellow (101, 7.67%), researcher (78, 5.92%), and other positions (e.g., director, medical doctor, research professor) (25, 1.90%). In regards to status, 790 participants (59.98%) received tenure, 187 participants (14.20%) are on tenure track, 268 participants (20.35%) are not on tenure track, 57 participants (4.33%) were retired, and 15 participants (1.14%) did not indicate their status. As for the education and work sector, all the participants (1,317, 100%) have PhD degrees and work in academic institutions. The summary of demographics of survey participants is presented in Table 1.

| | Demographic Category | Number | Percentage |
|---|---|---|---|
| Gender | Male | 936 | 71.07% |
| | Female | 348 | 26.42% |
| | *Missing* | 33 | 2.51% |
| Age | 25-34 | 139 | 10.55% |
| | 35-44 | 332 | 25.21% |
| | 45-54 | 334 | 25.36% |
| | 55-64 | 328 | 24.91% |
| | 65+ | 174 | 13.21% |
| | *Missing* | 10 | 0.76% |
| Ethnic | Asian/Pacific Islander | 167 | 12.68% |
| | Black/African-American | 14 | 1.06% |
| | Caucasian | 1,046 | 79.42% |
| | Hispanic | 32 | 2.43% |
| | Native American/Alaska Native | 1 | 0.08% |
| | Other/Multi-Racial | 27 | 2.05% |
| | *Missing* | 30 | 2.28% |
| Education | PhD/Doctoral Degree | 1,317 | 100.00% |
| Status | Tenured | 790 | 59.98% |
| | On Tenure Track | 187 | 14.20% |
| | Not On Tenure Track | 268 | 20.35% |
| | Retired | 57 | 4.33% |
| | *Missing* | 15 | 1.14% |
| Position | Lecturer/Instructor | 8 | 0.61% |
| | Professor of Practice | 6 | 0.46% |
| | Post-Doctoral Fellow | 101 | 7.67% |
| | Researcher | 78 | 5.92% |

| | | | |
|---|---|---|---|
| | Assistant Professor | 197 | 14.96% |
| | Associate Professor | 305 | 23.16% |
| | Full Professor | 544 | 41.31% |
| | Professor Emeritus | 53 | 4.02% |
| | Other | 25 | 1.90% |
| Sector | Academic | 1,317 | 100% |
| Total | | 1,317 | 100% |

Table 1. Demographics of Survey Participants

With regards to the academic disciplines, 1,317 survey participants belong to 43 STEM disciplines based on the NSF discipline codes. They are from seven disciplines of Engineering (181, 13.74%), three disciplines of Physical Sciences (93, 7.06%), three disciplines of Earth, Atmospheric, and Ocean Sciences (114, 8.66%), five disciplines of Agricultural Sciences (129, 9.79%), 14 disciplines of Biological Sciences (552, 41.91%), three disciplines of Psychology (77, 5.85%), five disciplines of Social Sciences (115, 8.73%), and three disciplines of Health Sciences (56, 4.25%). The discipline information of survey participants is shown in the Appendix B.

Scale Assessment

Scale assessment was conducted by using Cronbach's alpha and principal component factor analysis. Cronbach's alpha was used to estimate the internal consistency of multiple items for a construct and assess the extent to which a set of items belong to a construct. The Cronbach's alpha values in this research for each construct were greater than .70. They range from .867 for Regulative Pressure by Funding Agencies and Perceived Career Risk to .948 for Scholarly Altruism. In addition, each set of multiple measurement items for a construct was examined using item-total correlations to identify items that exhibited measurement problems or did not fit the core of each construct. All the items have item-total correlations ranging from .592 to .880, all of which are above the minimum recommended value of .50 (Field, 2009). Cronbach's alpha coefficients and item-to-total correlations are indicated in Table 2, saying that all the research constructs have satisfactory reliability values.

| Variable | Number of Items | Cronbach's alpha | Number of Cases Used | Item-to-Total Correlation |
|---|---|---|---|---|
| Regulative Pressure by Funding Agencies | 4 | .867 | 1210 | .646 - .800 |
| Regulative Pressure by Journal Publishers | 4 | .911 | 1177 | .739 - .859 |
| Normative Pressure by Disciplines | 4 | .875 | 1269 | .694 - .766 |
| Data Repository | 3 | .931 | 1251 | .846 - .878 |
| Perceived Career Benefit | 4 | .922 | 1273 | .734 - .876 |
| Perceived Career Risk | 4 | .867 | 1301 | .592 - .793 |
| Perceived Effort | 4 | .877 | 1277 | .710 - .766 |
| Scholarly Altruism | 6 | .948 | 1256 | .806 - .869 |

Table 2. Reliability Values (N=1,317)

The construct validity of the measurement items was assessed by using factor analysis. In this research, principal component factor analysis with Varimax rotation was performed by extracting factors with eigenvalues greater than 1. The results of factor analysis showed the existence of eight factors with eigenvalues greater than 1, and good convergent and discriminant validity. Together, the eight observed factors explained 78.13% of the total variance, which is considered satisfactory (Hair, Black, Babin, Anderson, & Tatham, 2006). All items loaded with factor loading value of .646 or more on each intended construct for which they were used to operationalize, showing good convergent validity. There were no cross-construct loadings above .293 for each factor, showing good discriminant validity. The results of scale assessment suggest that all the research constructs have satisfactory reliability and validity data.

## DATA ANALYSIS AND RESULTS

Consistent with the multilevel theoretical framework combining institutional theory (discipline level) and theory of planned behavior (individual level), a multilevel analysis was employed for this research, since the estimation of variances in different levels is theoretically relevant (Dansereau et al. 1995; Klein et al. 1994). The theoretical framework presented in this research shows that scientists' data sharing behaviors are expected to vary significantly, based on both on their discipline as well as individual factors. Individual scientists are nested within scientific disciplines, and this research assumes that the scientists in the same discipline share the same institutional influences. Variations in scientists' data sharing behaviors are partly attributable to scientists' perceptions and characteristics toward data sharing and partly attributable to the institutional influences in their disciplines. Multilevel analysis is an appropriate method for analyzing data in which one unit is nested within another higher level unit (Sacco et al. 2003).

This research employed a multilevel analysis method that examines the nested nature of social phenomena (e.g., students within schools) and accomplishes an integrated understanding of the multiple units of analysis. Among the diverse multilevel models by Kozlowski and Klein (2000), this research considered a cross-level direct-effect model. Such models examine how both higher-level predictors and lower-level predictors account for a lower-level outcome. In this research, the hierarchical data allowed a multilevel analysis with scientists nested within their disciplines. The multilevel analysis enabled examining the influence of both individual and discipline-level predictors on scientists' data sharing because it can simultaneously estimate the variation of scientists' data sharing behaviors based on individual and discipline-level predictors.

Before the multilevel analysis, the Intra-class Correlation Coefficients (ICCs) and $r_{wg(j)}$ statistics were used to assess whether the disciplinary-level variables were properly aggregated to the group level of analysis. Aggregated scales for discipline-level variables were created based on the individual scientists' responses on a set of items for each discipline-level construct. The individual responses for group level variables can be aggregated to the group level if there is a sufficient within-group agreement for considering group level variables as shared properties (Klein et al., 1994; Kozlowski & Klein, 2000). Therefore, it is important to check whether the aggregations of individual scientists' responses to the discipline-level variables are appropriate. The data aggregation statistics including ICC(1), ICC(2), and $r_{wg(j)}$ are indicated in Table 3.

| Group-Level Variable | ICC(1) | ICC(2) | $r_{wg(j)}$ |
|---|---|---|---|
| Regulative Pressure by Funding Agencies | 0.072 | 0.705 | 0.67 |
| Regulative Pressure by Journals | 0.182 | 0.872 | 0.65 |
| Normative Pressure | 0.086 | 0.742 | 0.76 |

| | | | |
|---|---|---|---|
| Data Repository | 0.156 | 0.850 | 0.70 |

Table 3. Data Aggregation Statistics for Discipline-Level Variables

The reliability statistics show that four discipline-level predictors can be aggregated to group level. ICC(1) values for regulative pressure by funding agencies (.072), regulative pressure by journal publishers (.182), normative pressure (.086), and repository (.156) were within the acceptable range (.05 to .20) (James, 1982), and ICC(2) values for regulative pressure by funding agencies (.705), regulative pressure by journal publishers (.872), normative pressure (.742), and repository (.850) were greater than .70, which is considered acceptable for data aggregation (Lindell, Brandt, & Whitney, 1999; Richardson & Vandenberg, 2005). Across the five discipline-level variables, the median $r_{wg(j)}$ values for normative pressure (.76) and data repository (.70) were above the .70 recommended value (Bliese, 2000; James, Demaree, & Wolf, 1993), and the median $r_{wg(j)}$ values for regulative pressure by funding agencies (.67) and regulative pressure by journal publishers (.65) were slightly below the .70 but still above the .60 acceptable value, suggesting moderate agreement (LeBreton & Senter, 2008). By considering any relevant indicators, discipline-level variables were aggregated to group level from the individual scientists' responses in each discipline.

The multilevel regression analysis in this research was performed using Hierarchical Linear Modeling (HLM) software. For the data analysis, the three-step multilevel modeling procedure (Hofmann, 1997) was conducted. First, the fully unconditional model with no individual and discipline level predictors was created, and this null model was used to determine what portions of the total variance in the dependent variable resided within and between groups. A one-way ANOVA was utilized to partition the variance in the dependent variable (data sharing behavior) within and between discipline components. This allowed determining whether there is significant between-discipline variance in scientists' data sharing behaviors (Raudenbush & Bryk, 2002).

*Unconditional Model*

The unconditional model (in which no discipline- and individual-level predictors were included other than scientists' data sharing behaviors) was formulated. Based on the unconditional model with one-way ANOVA, the between- and within- discipline variance in scientists' data sharing behaviors was estimated (Raudenbush & Bryk, 2002). The ANOVA results showed that there was significant between-discipline variance in scientists' data sharing behaviors ($F(1, 42) = 684.729$, $p<.001$). The $\chi^2$ test for the portion of variance in data sharing behaviors between disciplines was also significant ($\chi^2=352.065$, $p<.001$). This significant result suggests that further analysis for examining disciplinary-level influences on scientists' data sharing behaviors can be pursued using multilevel analyses. The results of these analyses were shown in Table 4.

| Fixed Effect | Coefficient | Standard Error | t-Ratio | P-Value |
|---|---|---|---|---|
| Data Sharing Behavior ($\gamma_{00}$) | 4.130 | 0.155 | 26.592 | <0.001 |
| **Random Effect** | **Variance Component** | ***df*** | **Chi-Square** | **P-Value** |
| Intercept ($u_0$) | 0.915 | 42 | 352.065 | <0.001 |
| Level 1 ($r$) | 3.865 | | | |

Table 4. Results from Unconditional Model

Based on the unconditional model, this research examined how much the amount of variance in scientists' data sharing behaviors resided within and between disciplines. The null model showed that the estimate for within-discipline (scientist level) variance was 3.865, and the between-discipline variance (discipline level) was 0.915. The Intraclass Correlation Coefficient (ICC) was calculated by the portion of disciplinary-level variance of the total variance, including disciplinary- and individual-level variances in the dependent variable (i.e., data sharing behavior) (Raudenbush & Bryk, 2002). The ICC for scientists' data sharing behaviors was .191 (0.915/(0.915+3.865)=.191), indicating that 19.1 percent of the total variance in scientists' data sharing behaviors existed between disciplines, while 80.9 percent of the variance existed within disciplines. In other words, the scientists' data sharing behaviors may vary between disciplines, and the scientists' data sharing behaviors were influenced by not only individual-level predictors, but also by discipline-level predictors.

*Individual Model*

The Level 1 model was estimated based on the individual-level variables only, with no discipline-level predictors included for the Level 2 model. The Level 1 model includes four individual-level variables (including perceived career benefit, perceived career risk, perceived effort, and scholarly altruism). The within-discipline variance ($\sigma^2$) has changed from 3.865 to 3.227, and this difference shows the portion of within-discipline variance explained by individual level predictors (Within-Group $R^2$=.165). These four individual-level independent variables explained 16.5 percent of the within-discipline variance ((3.865 – 3.227) / 3.865 = .165). After adding individual-level predictors, the residual variance at the disciplinary level ($\tau_{00}$) becomes low (from .915 to .588). This means that some of the between-discipline variance in data sharing behaviors was partially explained by those individual-level predictors identified in the Level 1 model. Table 5 below shows the results of the individual-level model as well as unconditional model and multilevel model.

| Predictors | | Step 1 Unconditional Model | Step 2 Individual-Level Predictors Only | Step 3 Adding Group-Level Predictors |
|---|---|---|---|---|
| Discipline Level Predictors | Funding Agencies' Pressure | | | -0.051 |
| | Journals' Pressure | | | $0.366^{**}$ |
| | Normative Pressure | | | $0.762^{**}$ |
| | Data Repository | | | 0.194 |
| | Residual Variance ($\tau_{00}$) | 0.915 | 0.588 | 0.129 |
| Individual Level Predictors | Perceived Career Benefit | | $0.088^{*}$ | $0.081^{*}$ |
| | Perceived Career Risk | | -0.010 | -0.008 |
| | Perceived Effort | | $-0.142^{***}$ | $-0.138^{***}$ |
| | Scholarly Altruism | | $0.688^{***}$ | $0.667^{***}$ |
| | Residual Variance ($\sigma^2$) | 3.865 | 3.227 | 3.229 |
| Within-Group $R^2$ | | | 0.165 | |
| Between-Group $R^2$ | | | | 0.781 |

$^{***}p$<.001, $^{**}p$<.01, $^{*}p$<.05

Table 5. Fixed-Effect Results for Data Sharing Behavior

*Multilevel Model*

The multilevel model was estimated by using both Level 1 and Level 2 predictors. Based on the Level 1 model, four discipline-level predictors (including funding agencies' regulative pressure, journals' regulative pressure, normative pressure, and data repository) were added into the multilevel model. The between-discipline variance ($\tau_{00}$) has changed from 0.588 to 0.129, and this difference shows the portion of between-discipline variance explained by discipline-level predictors (Between-Group $R^2$=.781). These four discipline-level predictors accounted for 78.1 percent of the between-discipline variance in data sharing behaviors ((0.588 – 0.129) / 0.588 = .781). Table 6 below shows the results of multilevel model including unstandardized beta and standard error, t-value, and p-value.

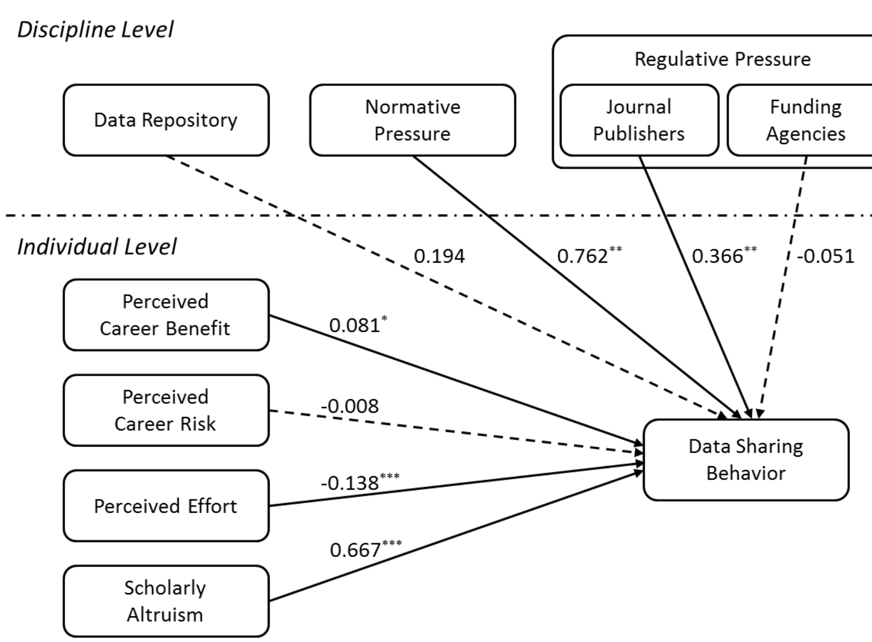
| Fixed Effect | Unstandardized Coefficients | | Standardized Beta | t-ratio | *p*-value |
|---|---|---|---|---|---|
| | Beta | Std. Error | | | |
| *Discipline Level* | | | | | |
| Funding Agencies' Pressure | -0.051 | 0.243 | -0.012 | -0.210 | 0.835 |
| Journals' Pressure | 0.366 | 0.130 | 0.140 | 2.826 | 0.007 |
| Normative Pressure | 0.762 | 0.216 | 0.184 | 3.526 | 0.001 |
| Data Repository | 0.194 | 0.148 | 0.064 | 1.311 | 0.198 |
| *Individual Level* | | | | | |
| Perceived Career Benefit | 0.081 | 0.037 | 0.059 | 2.179 | 0.030 |
| Perceived Career Risk | -0.008 | 0.041 | -0.006 | -0.203 | 0.839 |
| Perceived Effort | -0.138 | 0.041 | -0.085 | -3.368 | <0.001 |
| Scholarly Altruism | 0.667 | 0.060 | 0.315 | 11.081 | <0.001 |

| Random Effect | Variance Component | *df* | Chi-Square | p-value |
|---|---|---|---|---|
| Intercept | 0.129 | 38 | 81.199 | <0.001 |
| Level 1 | 3.229 | | | |

Table 6. Results from Research Model (2 Level Model)

The data analysis results showed how both individual and institutional factors influence scientists' data sharing behaviors at the same time. At the individual level, perceived career benefit (β=0.081, p<0.05) and scholarly altruism (β=0.667, p<.001) are found to have significant positive relationships with scientists' data sharing behaviors, and perceived effort (β=-0.138, p<.001) is found to have a significant negative relationship with scientists' data sharing behaviors. Perceived career risk (β=-0.008, p=0.839), however, is not found to be significantly related to scientists' data sharing behaviors.

At the discipline level, both regulative pressure by journals (β=0.366, p<0.01) and normative pressure (β=0.762, p<0.01) are found to have significant positive relationships with data sharing behaviors; however, regulative pressure by funding agencies (β=-0.051, p=0.835) is not found to have a significant relationship with data sharing behaviors. Also, the availability of data repositories (β=0.194, p=0.198) is not found to be significantly related to scientists' data sharing behaviors. Overall $R^2$ based on both discipline and individual level predictors was 0.297. By separating the overall $R^2$ into between-discipline and within-discipline portions, 16.4 percent and 13.3 percent of total variances in scientists' data sharing behaviors are explained by the discipline-level and individual-level predictors respectively. Therefore, this research demonstrated that scientists' data sharing behaviors are influenced by both institutional factors (i.e., regulative pressure by journals and normative pressure) and individual factors (i.e., perceived career benefit, perceived effort, and scholarly altruism). Figure 2 below shows the summary of hypothesis testing results:



Unstandardized Beta, $^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$

Figure 2. Hypothesis Testing Results based on Scientists' Data Sharing Behavior Model

## DISCUSSIONS

In this section, we provided detailed discussions of the research findings based on the multilevel analysis. We also used some excerpts from the responses for the open-ended question asking survey participants to "share any additional comments, questions, or suggestions about scientific data sharing." Those excerpts were included in the discussions in order to help us to interpret the survey results.

### *Individual Level Predictors*

#### Perceived Career Benefit

Perceived career benefit was found to have a significant positive influence on scientists' data sharing behaviors. This means that scientists who perceive there are more career benefits in sharing data in their published articles are more likely to share their data with others. This result supports prior studies' findings that professional recognition (Kim, 2007), institutional recognition (Kankanhalli et al., 2005), and academic reward (Kling & Spector, 2003) all influence scientists' data sharing behaviors. Recognition and reputation through increased citations and possible credits are associated with the concept of perceived career benefits. This research shows that in the perspective of motivation, scientists' data sharing behaviors are driven by their perceived values of their behaviors and by the rewards they expect to derive from sharing their data. According to Merton's (1957) value of science, the motivation for scientists to achieve reputation is an important value in many scientific communities. The finding of this research suggests that scientists' motivation to achieve academic recognition and credit can increase their data sharing behaviors, and we can encourage scientists' data sharing behaviors by better utilizing the reward system in scientific communities.

Prior studies in knowledge sharing also found that expected social rewards from knowledge sharing behavior have a positive effect on individuals' attitudes toward knowledge sharing and their intentions to share knowledge (Hsu & Lin, 2008; Jones, Hesterly, & Borgatti, 1997; Kim & Han, 2009). The concept of reward through recognition and reputation is a well-known factor influencing knowledge sharing behavior (Hung, Lai, & Chang, 2011b). This research shows that in the context of scientists' data sharing, as scientists perceive more career benefits through recognition and reputation, they are more willing to share their data with others. This finding is also related to Piwowar and colleagues' (2007) finding that articles that provided their relevant data sets (i.e., microarray data) through data repositories received more citations than articles that did not provide their data sets.

Perceived Career Risk

In this research, perceived career risk was not found to have a significant relationship with scientists' data sharing behaviors. Prior studies argued that scientists view data sharing as potential loss (e.g., losing publication opportunities) or impediment for their careers, so they are reluctant to share their data (Louis et al., 2002; Reidpath & Allotey, 2001; Savage & Vickers, 2009; Stanley & Stanley, 1988). However, this research did not find any significant negative relationship between perceived career risk and scientists' data sharing behaviors. One possible reason for this insignificant result is that data sharing in this research is conceptualized as sharing the data of published articles only rather than the data of unpublished articles. Therefore, the different concepts of data sharing in each research need to be considered in interpreting this finding.

Scientists have concerns about sharing the data of unpublished work, but they are less concerned about sharing the data of published articles. Several survey participants provided the comments that they are less concerned about sharing the data of published articles. A scholar in plant science mentioned, "I avoid sharing sensitive data before it is published because I do not want my students and postdocs to be scooped. [...] Once we are published, then we share our data and the scientific materials with any who want them." Therefore, this research suggests that perceived career risk involved in sharing the data of published articles does not have a significant negative effect on scientists' data sharing behaviors (i.e., sharing the data of published articles).

## Perceived Effort

Perceived effort was found to have a significant negative effect on scientists' data sharing behaviors. This means that scientists who perceive that it requires more effort to participate in data sharing are less likely to share their data with others. This result supports many of prior studies' arguments that the efforts (e.g., additional work, cost, and time) involved in data sharing discourage scientists to share their data (Campbell et al., 2002; Foster & Gibbons, 2005; Louis et al., 2002; Tenopir et al., 2011). This finding is also relevant to what Tenopir and colleagues (2011) recently found: scientists do not make their data available online because they lack the time and funding to organize their data.

Data sharing requires a lot of time and effort from scientists to make their data accessible. Scientists need to organize and arrange their data sets for other scientists, and sometimes they also need to provide extensive explanations about their data in order to help other scientists make sense of the data sets. Therefore, many scientists have concerns about the efforts involved in data sharing, so perceived effort negatively influences scientists' data sharing behaviors. A scholar in electrical engineering emphasized the issue of extra effort required in data sharing, saying: "For many small experiments, the amount of effort required to fully organize, document, and explain data to an outside researcher is greater than the effort required to simply recreate the experiment."

## Scholarly Altruism

Scholarly altruism was found to have a significant relationship with scientists' data sharing behaviors. This finding agrees with prior studies' findings that altruism has a significant influence on information sharing behaviors (Hsu & Lin, 2008). In the context of data sharing, a few prior studies discovered that altruism is an important factor influencing faculty members' contribution to institutional data repositories (Foster & Gibbons, 2005; Kim, 2007); in the context of knowledge sharing, altruism was extensively studied and found to have significant influence on knowledge sharing (Constant, Sproull, & Kiesler, 1996;

Davenport, & Prusak, 1998; He & Wei, 2009; Hung, Durcikova, Lai, & Lin, 2011a; Kankanhalli et al., 2005; Lin, 2008).

Some of previous studies in information sharing defined the concept of altruism as a form of intrinsic motivation (i.e., having psychological benefits such as satisfaction and enjoyment of helping others) (Cho, Chen, & Chung, 2010; Hung et al., 2011a; Hung et al., 2011b; Lee & Lee, 2010); however, this research redefines "scholarly altruism" by focusing on individual's willingness to work to increase others' welfare and contribute to their communities without expecting anything in return (Hsu & Lin, 2008). This research shows that scholarly altruism motivates scientists to help other scientists to save time and effort, allowing them to find something missing from the original research, and contributing to scientific development in their fields through data sharing.

*Institutional Level Predictors*

Regulative Pressure by Funding Agencies

Regulative pressure by funding agencies was not found to have a significant relationship with scientists' data sharing behaviors, and this finding is different from what prior research argued. Prior studies found that data sharing policies by funding agencies have positive influences on scientists' data sharing (McCullough et al., 2008; Piwowar & Chapman, 2008); however, this research did not find a significant correlation between regulative pressure by funding agencies and scientists' data sharing behaviors. The discrepancy of the findings between prior studies and this research may be resulting from the differences in disciplines included for each research. Prior studies focused on certain disciplines in biological sciences (Piwowar, 2011; Piwowar & Chapman, 2008); however, this research extended to diverse STEM disciplines.

Many scholars argued that funding agencies' data sharing policies would increase scientists' data sharing behaviors (McCullough et al., 2008; Piwowar & Chapman, 2008; Stanley & Stanley, 1988); however, this research did not find a positive correlation between funding agencies' regulative pressure and scientists' data sharing behaviors across diverse STEM disciplines. One possible interpretation of this non-significant result is that since the data management requirement with strong encouragement for data sharing by NSF was implemented fairly recently (National Science Foundation, 2010), the effects of funding agencies' push was not reflected in scientists' data sharing behaviors yet. The analysis of preliminary interviews showed that there were two different perspectives regarding NSF's new data sharing policy. A professor in environmental engineering mentioned:

"Every proposal has a data sharing policy now. And so we were rewarded, and I mean, I guess we are penalized for not sharing data because you won't get your grant unless you have a policy for sharing your data. So I think that you know the question about not sharing data is now moot because NSF funded most of our research. We have to share our data."

However, another professor in biology mentioned that the NSF policy did not have a significant impact on scientists' data sharing, by saying:

"I haven't seen much of it yet, how NSF's changes [of data management policy] will affect people because it's a relative new requirement. […] And NSF themselves, I was personally at NSF when they were making these changes, and even then, program officers at NSF weren't taking it particularly seriously. […] So, you know, if it meant the difference between your proposal being funded and not being funded, then people are going to take it very seriously. But it was just an extra thing you had to write."

In addition, it also might be possible that scientists do not perceive funding agencies' data sharing policies as a serious coercive pressure, even if the agencies have had data sharing policies for a while (e.g., biological and health sciences funded by NIH). A number of survey participants commented that national funding agencies do not enforce their data sharing policies, so scientists do not perceive any serious coercive pressures from funding agencies. A professor in neuroscience mentioned:

"There is little institutional/funding pressure to do so [data sharing]. NIH (biomedical funding) requires data sharing, but [it is] only taken seriously by a few disciplines (genomic data, brain imaging). As far as I can tell there are no explicit checks on whether data sharing occurs or penalties if the data [are] is not made available."

This shows that although there are data sharing policies required by funding agencies (NSF and NIH), scientists do not perceive any serious coercive pressures from those policies because (1) the data sharing policies were implemented recently (i.e., NSF), and (2) funding agencies do not explicitly enforce their data sharing policies except particular discipline(s) (i.e., NIH). Therefore, it can be concluded that regulative pressure by funding agencies does not have a significant influence on scientists' data sharing behavior across diverse STEM disciplines.

Regulative Pressure by Journals

This research found that journals' regulative pressure has a significant influence on scientists' data sharing behaviors. This finding demonstrates that journals exert strong coercive pressures on scientists' data sharing behaviors. This finding is consistent with some of the prior bibliometric studies' findings that there are positive correlations between the existence of data sharing policy in journals and the rate at which scientists deposit data in public databases (Piwowar & Chapman, 2008; Piwowar & Chapman, 2010). However, other studies argued that the data sharing policies in certain journals did not have significant impacts on actual data sharing rates (Cech et al., 2003).

Compared to prior studies, this research examined the relationship between regulative pressure by journals and scientists' data sharing behaviors across different science and engineering disciplines, and found that regulative pressure by journals in each discipline positively increases scientists' data sharing behaviors. A good number of journals in biological sciences have required their authors to submit data either as supplements or in data repositories as a condition of publication, and more journals (e.g., evolutionary biology and ecology) recently have implemented data sharing policies that require their authors to share data by depositing it into data repositories (Savage & Vickers, 2009; Weber, Piwowar, &

Vision, 2010). This research shows that there is a significant relationship between the regulative pressure by journals in each discipline and scientists' data sharing behaviors.

Normative Pressure

This research found that normative pressure from each scientific discipline (or community) significantly influence scientists' data sharing behaviors across different disciplines. Prior studies did not examine the relationship between the normative pressure in each discipline and their scientists' data sharing behaviors as yet. This research showed that there are significant between-discipline variances in normative pressure, and normative pressure in each discipline positively influences scientists' data sharing behaviors. This finding supports the idea that the scientific community's consensus toward data sharing is critical to facilitate scientists' data sharing behaviors (Zimmerman, 2007). With regards to Merton's (1973) norm of science, this finding suggests that scientists regard data sharing as a valuable norm of science. Especially, in terms of communalism, scientists would consider that scientific data related to their findings need to be shared with their communities, and this norm of science incorporated in normative pressure would positively increase scientists' data sharing behaviors.

The normative pressures can be formulated as the forms of professionalism and expectation from peer-scientists in a scientific community. Scientists need to conform to the established norms in their disciplines in order to maintain their legitimacy and conduct research with other scientists. This research shows normative pressures differ across diverse scientific disciplines, and normative pressure plays an important role in scientists' data sharing behaviors. Scientists socially agree on their data sharing practices and follow the socially adopted norms about their data sharing. Therefore, scientists in the disciplines that have strong normative pressures about data sharing are more likely to share their data with other scientists, In other words, scientists in the disciplines with low normative pressures are less likely to share their data.

Data Repository

The availability of data repositories in a discipline was not found to have a significant relationship with scientists' data sharing behaviors. Although a prior study showed that the lack of data repositories was an important barrier for data sharing in several disciplines (Kim & Stanton, 2012), this survey study did not confirm the positive relationship between the availability of data repositories in each discipline and scientists' data sharing behaviors. Prior studies argued that the existences of data repositories facilitate and promote scientists' data sharing in certain disciplines (e.g., molecular biology) (Brown, 2003; Cragin et al., 2010; Marcial & Hemminger, 2010). However, this research examined the relationship between the availability of data repositories and data sharing behaviors across diverse scientific disciplines, and it did not find any significant relationship.

This result showed that the availability of data repositories does not necessarily increase scientists' data sharing behaviors. The comments provided by survey participants indicate that the existing data repositories in some disciplines do not support scientists' data sharing due to the difficulties and the lack of supports in using those repositories. A microbiologist mentioned that, "NCBI Pubmed is a data repository that is so onerous to submit to (e.g., multiple genomes), that there is a significant barrier to data fidelity in this important public repository." Also, the existing data repositories in each discipline do not allow scientists to share all types of data generated in their disciplines. Another scholar in psychology mentioned that, "In my sub-field, there is one prominent and well respected repository for sharing raw data -- it's the CHILDES website. But this is a place for naturalistic data, not experimental work. While it is some trouble to post to CHILDES (formatting, permissions, etc.) it is well respected." Although this finding seems unexpected, the availability of data repositories in each discipline may provide some explanation for scientists' data sharing behaviors.

## Limitations of the Study

This research has a few limitations that should temper interpretation and application of the results. First, the recruiting method employed in the survey study may have resulted in self-selection bias or nonresponse bias. Although the sampling frame was randomly selected from the CoS scholar database, the field survey ultimately involved those participants who voluntarily participated in the survey. The overall response rate was only 15.28%, leaving ample opportunity for concerns about either kind of bias. According to a meta-analysis of the response rates for online surveys in the articles published from 1995 to 2006, the average response rate was only 34% (Shih and Fan, 2008). Another recent study by Allen (2010) found that the average response rate has been decreased significantly by 24% based on the surveys conducted between 2005 and 2010. Although the response rates for online surveys are low, it is necessary to increase the overall response rate in order to prevent any nonresponse bias. The future research can employ online and mail surveys together as suggested by Millar and Dillman (2011), since both email and mail addresses of scientists are available on the CoS database.

Another methodological limitation of survey was the self-report nature of the dependent measures. The survey method included self-reports regarding the measurement of scientists' data sharing behaviors. Each participant was asked to provide their own data sharing behaviors: Prior research has shown that individuals often respond to surveys in a way that presents a positive impression, so scientists may have over-reported their actual data sharing. It was not feasible for us to obtain more objective observations of their actual behaviors in this research project. In addition, the survey participants might interpret data sharing differently in spite of the definition of data sharing was provided at the beginning of the survey. For example, some scientists in a certain discipline may consider research materials (e.g., reagent) and protocols (e.g., source codes) as raw data to be shared; however, other scientists in the same discipline may not consider those materials and protocols as a part of raw data.

The multilevel method utilized in this research also has limitations. One of the limitations is that the discipline-level constructs may have a potential bias in their measurements. This research measured the discipline-level constructs by aggregating individual scientists' reports about their discipline-level information. This may not measure the exact status of group-level constructs, and it may cause a potential bias in group-level measurements. In a different perspective, it is possible to consider the institutional factors as individual-level constructs by assuming that scientists' perceptions toward institutional factors would differ within each discipline and consequently influence scientists' data sharing behaviors. Therefore, based on the same dataset, we can validate the research model by focusing on each major discipline (e.g., biological science) and further conduct a comparison study with different major disciplines (e.g., biological science vs. physical science).

Finally, another limitation of the multilevel method in this research is the small group size for several disciplines included in the final analysis. Although at least 20 observations in one group are recommended by recent studies (Hox, 2002; Scherbaum & Ferreter, 2009), this research included five disciplines (out of forty-three disciplines) that contained fewer than 20 members (but still more than 15 members) for its multilevel analysis. The small group sizes for those five disciplines may have a potential problem with internal consistency; however, we decided to include those five disciplines in order to increase the statistical power to detect the discipline-level (Level 2) predictors.

## CONCLUSION

This research investigated how both institutional environments and individual motivations influence scientists' data sharing behaviors across diverse disciplines. The results of this research showed that both institutional pressures (i.e., regulative pressure by journals and normative pressure in disciplines) and individual motivations (i.e., perceived career benefit, perceived effort, and scholarly altruism) have significant relationships with scientists' data sharing behaviors. The findings of this research suggested that in order to encourage data sharing, we need to consider both institutional environments and individual motivations simultaneously.

The research findings have several theoretical implications for institutional theory and theory of planned behavior. First, the integration of institutional theory and individual motivation theory provided a new theoretical lens to understanding scientists' data sharing behaviors. The combined theoretical framework was found to nicely account for how institutional and individual factors influence scientists' data sharing behaviors simultaneously. Second, with regards to institutional theory, this study sheds light on how institutional environments can influence individuals' behaviors. The results of this research showed the micro-foundations of institutions by looking at institutional influences and individual motivations together. Consequently, this research can advance the neo-institutional theory by applying it to the individual levels. Third, the research also contributed to the theory of planned behavior. The results of this study showed that perceived career benefit and perceived effort have direct relationships with actual data sharing behaviors. Those results support prior studies looking at the direct relationships between perceptions and actual behaviors based on the theory of planned behavior (Shi, Shambare, & Wang, 2008; Watson & Hewett, 2006; W.-L. Wu, Lin, Hsu, & Yeh, 2009).

This research also has methodological implications. This research utilized a multilevel analysis method in order to incorporate the multilevel theoretical framework and analyze the hierarchical data (i.e., scientists nested within their disciplines). Prior studies have predominantly examined scientists' data sharing as an individual phenomenon ignoring its institutional context; however, it is important to examine institutional influences as well as individual motivations together in studying scientists' data sharing behaviors. The multilevel regression analysis was employed to validate the multilevel research model, which was developed based on institutional theory and theory of planned behavior. Another methodological contribution of this research is the scale development procedure, taken to develop the measurement items to be used in the context of scientists' data sharing. Since the existing measurement items were not applied and tested in scientists' data sharing contexts, and there were potential gaps between existing items and constructs studied in this research, it was necessary to develop a dedicated measurement scale for studying scientists' data sharing behaviors. This research systematically developed its scales by validating the existing measurement items and creating new measurement items for its research model.

This research also proposes practical implications. Scientific data sharing can be promoted by the joint efforts of funding agencies, journal publishers, professional associations, and research institutions. This research argues that the vision of scientific data sharing can be achieved through (1) implementing funding agencies' and journals' data sharing policies with strong enforcement, (2) building community norms of data sharing through education and promotion supported by professional associations, (3) developing a good incentive system to provide appropriate credits for data sharing, (4) reducing the efforts involved in data sharing by standardizing data sharing protocols and providing data curation and management supports, and (5) lastly, facilitating individual scientists' scholarly altruism by creating an altruistic culture of data sharing in a scientific community.

This research showed a holistic picture of the phenomena of scientific data sharing across diverse disciplines rather than focusing on a particular case of data sharing in a discipline. Scientific data sharing practices may differ across disciplines. Even in disciplines where scientists generate different types of data, each discipline may have different data sharing requirements and expectations, which would be perceived differently by the group members. Therefore, future research also needs to investigate how data sharing factors and its methods differ across different disciplines and positions (e.g., student researchers vs. faculty members), and what contribute to those differences. Furthermore, the future studies need to address how both discipline-level and individual-level factors influence different forms of data sharing respectively. Since scientists' data sharing factors are influenced by the contexts of data sharing, it is important to investigate how both discipline- and individual-level factors interact with each other toward different forms of data sharing behaviors. The cluster analysis with data sharing factors can be a possible solution for understanding how data sharing factors are related to each other under different data sharing circumstances.

Also, future research needs to further examine some of the research constructs studied in this research; for example, scholarly altruism can be considered as a discipline-level construct showing a professional norm in each discipline, and discipline-level constructs can be considered as individual-level constructs in the perspective that institutional pressures can be perceived differently by individual scientists. Moreover, future research also needs to consider data reuse issues along with data sharing. This series of research endeavors can help us better understand scientists' data sharing behaviors. The findings of those research efforts can accelerate scientific collaborations and eventually advance scientific development in diverse scientific disciplines.

Acknowledgements

We would like to acknowledge the CoS Pivot (ProQuest) for allowing us to use their scholar database in recruiting the survey participants.

## REFERENCES


Allen, T. (2010). Student Surveys Conducted at SESRC (2005–2010). Unpublished data from the Social and Economic Sciences Research Center. Washington State University, Pullman, WA.

Ajzen, I. (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Process, 52(2), 179-211.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., ... Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. Data Science Journal, 3(0), 135-152.

Baytiyeh, H., & Pfaffman, J. (2010). Open source software: A community of altruists. Computers in Human Behavior, 26(6), 1345-1354.

Bietz, M. J., Baumer, E. P. S., & Lee, C. P. (2010). Synergizing in cyberinfrastructure development. Computer Supported Cooperative Work: CSCW: An International Journal, 19(3-4), 245-281.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis, in Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions, K. J. Klein & S. W. l. Kozlowski (eds.), San Francisco, CA: Jossey Bass, Inc.

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(1), 87-111.

Borgman, C. L. (2007). Scholarship in the digital age: Information, infrastructure, and the internet. Cambridge: MIT Press.

Borgman, C. L. (2010). Research Data: Who will share what, with whom, when, and why? Paper presented at the Fifth China - North America Library Conference, Beijing, China.

Brown, C. (2003). The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance. Journal of the American Society for Information Science and Technology, 54(10), 926-938.

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(2), 241-255.

Campbell, E. G., Clarridge, B. R., Gokhale, N. N., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics - Evidence from a national survey. Jama-Journal of the American Medical Association, 287(4), 473-480.

Campbell, E. G., Louis, K. S., & Blumenthal, D. (1998). Looking a gift horse in the mouth - Corporate gifts supporting life sciences research. Jama-Journal of the American Medical Association, 279(13), 995-999.

Campbell, E. G., Weissman, J. S., Causino, N., & Blumenthal, D. (2000). Data withholding in academic medicine: characteristics of faculty denied access to research results and biomaterials. Research Policy, 29(2), 303-312.

Cech, T. R., Eddy, S. R., Eisenberg, D., Hersey, K., Holtzman, S. H., Poste, G. H., . . . Authorship, B. (2003). Sharing publication-related data and materials: Responsibilities of authorship in the life sciences. Plant Physiology, 132(1), 19-24.

Ceci, S. J. (1988). Scientists Attitudes toward Data Sharing. Science Technology & Human Values, 13(1-2), 45-52.

Cho, Hichang, Chen, MeiHui, & Chung, Siyoung. (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(6), 1198-1212.

Choudhury, G. S. (2008). Case Study in Data Curation at Johns Hopkins University. Library Trends, 57(2), 211-220.

Cohen, J. (1995). Share and Share alike isn't Always the Rule in Science. Science, 269(5227), 1120-1120.

Constant, D., Sproull, L., & Kiesler, S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. Organization Science, 7(2), 119-135.

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 368(1926), 4023-4038.

Cronin, B. (1984). The Citation Process: The Role and Significance of Citations in Scientific Communication. London: Taylor Graham.

Cronin, B. (2005). The hand of science: academic writing and its rewards. Oxford, UK: Scarecrow Press, Inc.

Dansereau, F., Yammarino, F. J., & Markham, S. E. (1995). Leadership: The multiple-level approaches. The Leadership Quarterly, 6(2), 97-109.

Davenport, T. H., & Prusak, L. (1998). Working knowledge: How organizations manage what they know. Boston, MA: Harvard Business School Press.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance in Information Technology. MIS Quarterly, 13(3), 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(8), 982-1003.

Davis, H. M., & Vickery, J. N. (2007). Datasets, a shift in the currency of scholarly communication: Implications for library collections and acquisitions. Serials Review, 33(1), 26-32.

DiMaggio, P. J., & Powell, W. W. (1991). Introduction. In W. W. Powell & P. J. DiMaggio (Eds.), The New Institutionalism in Organizational Analysis (pp. 1-38). Chicago: The University of Chicago Press.: The University of Chicago Press.

Fabrigar, Leandre R, Wegener, Duane T, MacCallum, Robert C, & Strahan, Erin J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological methods, 4(3), 272-299.

Faniel, Ixchel M., & Zimmerman, Ann. (2011). Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse. International Journal of Digital Curation, 6(1), 58-69.

Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. International Journal of Human Computer Studies, 59(4), 451-474.

Field, A. (2009). Discovering Statistics Using SPSS (3rd ed.). Thousand Oaks, CA: Sage Publications.

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior. Reading, MA: Addison-Wesley.

Foster, N. F., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. D-Lib Magazine, 11(1).

Goldstein, Harvey. (2011). Multilevel Statistical Models (4th ed.). West Sussex, UK: John Wiley & Sons, Ltd.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis (6th ed.). Upper Saddle River, NJ: Prentice Hall.

He, Wei, & Wei, Kwok-Kee. (2009). What drives continued knowledge sharing? An investigation of knowledge-contribution and -seeking beliefs. Decision Support Systems, 46(4), 826-838.

Hey, T., & Trefethen, A. E. (2004). UK e-Science Programme: Next generation grid applications. International Journal of High Performance Computing Applications, 18(3), 285-291.

Hey, T., & Trefethen, A. E. (2008). E-science, cyberinfrastructure, and scholarly communication. In G. M. Olson, A. Zimmerman & N. Bos (Eds.), Scientific collaboration on the Internet. Cambridge, MA: MIT Press.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. Journal of Management, 23(6), 723-744.

Hox, J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Hsu, Chin-Lung, & Lin, Judy Chuan-Chuan. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. Information & Management, 45(1), 65-74.

Hung, Shin-Yuan, Durcikova, Alexandra, Lai, Hui-Min, & Lin, Wan-Mei. (2011a). The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. International Journal of Human-Computer Studies, 69(6), 415-427.

Hung, Shin-Yuan, Lai, Hui-Min, & Chang, Wen-Wen. (2011b). Knowledge-sharing motivations affecting R&D employees' acceptance of electronic knowledge repository. Behaviour & Information Technology, 30(2), 213-230.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. Journal of Applied Psychology, 67(2), 219-229.

James, L. R., Demaree, R. G., & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. Journal of Applied Psychology, 78(2), 306-309.

John, C. H. S., Cannon, A. R., & Pouder, R. W. (2001). Change drivers in the new millennium: implications for manufacturing strategy research. Journal of Operations Management, 19(2), 143-160.

Jones, C., Hesterly, W. S., & Borgatti, S. P. (1997). A general theory of network governance: Exchange conditions and social mechanisms. Academy of Management Review, 22(4), 911-945.

Kankanhalli, A., Tan, B. C. Y., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. MIS Quarterly, 29(1), 113-143.

Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure Time: Long-term Matters in Collaborative Development. Computer Supported Cooperative Work-the Journal of Collaborative Computing, 19(3-4), 377-415.

Kellogg, D. (2006). Toward a Post-Academic Science Policy: Scientific Communication and the Collapse of the Mertonian Norms. International Journal of Communications Law and Policy, Fall 2006.

Kim, Byoungsoo, & Han, Ingoo. (2009). The role of trust belief and its antecedents in a community-driven knowledge environment. Journal of the American Society for Information Science and Technology, 60(5), 1012-1026.

Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. Journal of Digital Information, 8(2).

Kim, Youngseek, & Stanton, Jeffrey M. (2012). Institutional and individual influences on scientists' data sharing practices. Journal of Computational Science Education, 3(1), 47-56.

Klein, K. J., Dansereau, F., and Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. Academy of Management Review, 19(2), 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(3), 211-236.

Kling, R., & Spector, L. (2003). Rewards for scholarly communication. In D. L. Andersen (Ed.), Digital scholarship in the tenure, promotion, and review process. Armonk, NY: M.E. Sharpe, Inc.

Kostova, T., & Roth, K. (2002). Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects. Academy of Management Journal, 45(1), 215-233.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. l. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions. San Francisco, CA: Jossey Bass, Inc.

Lawrence, T., Suddaby, R., & Leca, B. (2011). Institutional Work: Refocusing Institutional Studies of Organization. Journal of Management Inquiry, 20(1), 52-58.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. Organizational Research Methods, 11(4), 815-852.

Lee, Gyudong, & Lee, Won Jun. (2010). Altruistic traits and organizational conditions in helping online. Computers in Human Behavior, 26(6), 1574-1580.

Lin, Chieh-Peng. (2008). Clarifying the relationship between organizational citizenship behaviors, gender, and knowledge sharing in workplace organizations in Taiwan. Journal of Business and Psychology, 22(3), 241-250.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. Applied Psychological Measurement, 23(2), 127-135.

Liu, H., Ke, W., Wei, K. K., Gu, J., & Chen, H. (2010). The role of institutional pressures and organizational culture in the firm's intention to adopt internet-enabled supply chain management systems. Journal of Operations Management, 28(5), 372-384.

Louis, K. S., Jones, L. M., & Campbell, E. G. (2002). Sharing in science. American Scientist, 90(4), 304-307.

Marcial, L. H., & Hemminger, B. M. (2010). Scientific Data Repositories on the Web: An Initial Survey. Journal of the American Society for Information Science and Technology, 61(10), 2029-2048.

McCain, K. W. (1995). Mandating sharing - Journal policies in the natural-sciences. Science Communication, 16(4), 403-431.

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? Canadian Journal of Economics-Revue Canadienne D Economique, 41(4), 1406-1420.

Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. Chicago: University of Chicago Press.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. Public Opinion Quarterly, 75(2), 249-269.

Mitroff, I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. American Sociological Review, 39, 569-595.

Mulkay, M. J. (1976). Norms and ideology in science. Social Science Information, 15(4-5), 637-656.

National Science Foundation. (2010). Scientists seeking NSF funding will soon be required to submit data management plans. from http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce, 7(3), 101-134.

Pfeffer, J., & Salancik, G. (1978). External Control of Organizations: A Resource Dependence Perspective. New York: Harper and Row.

Phang, Chee Wei, Sutanto, Juliana, Kankanhalli, Atreyi, Li, Yan, Tan, Bernard C.Y., & Teo, Hock-Hai. (2006). Senior Citizens' Acceptance of Information Systems: A Study in the Context of e-Government Services. IEEE Transaction on Engineering Management, 53(4), 555-569.

Piwowar, H. A. (2011). Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. Plos One, 6(7).

Piwowar, H. A., & Chapman, W. W. (2008). A review of journal policies for sharing research data. Paper presented at the International Conference on Electronic Publishing, Toronto Canada.

Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: A pilot study of associations. Journal of Informetrics, 4(2), 148-156.

Piwowar, H. A., Day, R. S., and Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. Plos One, 2(3), 1-5.

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(3), 17-82.

Rakov, T., & Marcoulides, G. A. (2000). A First Course in Structural Equation Modeling. Mahwah, NJ: Lawrence Erlbaum Associates.

Raudenbush, Stephen W, & Bryk, Anthony S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (2nd Ed.). Thousand Oaks: Sage Publications.

Reidpath, D. D., & Allotey, P. A. (2001). Data sharing in medical research: An empirical investigation. Bioethics, 15(2), 125-134.

Richardson, H. A., & Vandenberg, R. J. (2005). Integrating managerial perceptions and transformational leadership into a work-unit level model of employee involvement. Journal of Organizational Behavior, 26(5), 561-589.

Ryu, S., Ho, S. H., & Han, I. (2003). Knowledge sharing behavior of physicians in hospitals. Expert Systems with Applications, 25(1), 113-122.

Savage, C. J., & Vickers, A. J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. Plos One, 4(9).

Scherbaum, Charles A, & Ferreter, Jennifer M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. Organizational Research Methods, 12(2), 347-367.

Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. Journal of Applied Psychology, 88(5), 852-865.

Scott, R. W. (2001). Institutions and Organizations, 2nd Edition. Thousand Oaks, CA: Sage Publications.

Shi, W., Shambare, N., & Wang, J. (2008). The adoption of internet banking: An institutional theory perspective. Journal of Financial Services Marketing, 12(4), 272-286.

Shih, T.-H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. Field methods, 20(3), 249-271.

So, J., & Bolloju, N. (2005). Explaining the intentions to share and reuse knowledge in the context of IT service operations. Journal of Knowledge Management, 9(6), 30-41.

Son, J. Y., & Benbasat, I. (2007). Organizational Buyers' adoption and use of B2B electronic marketplaces: Efficiency- and legitimacy-oriented perspectives. Journal of Management Information Systems, 24(1), 55-99.

Stanley, B., & Stanley, M. (1988). Data sharing. The primary researcher's perspective. Law and Human Behavior, 12(2), 173-180.

Sterling, T. D., & Weinkam, J. J. (1990). Sharing scientific-data. Communications of the ACM, 33(8), 112-119.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. Information Systems Research, 6(2), 144-176.

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(6).

Teo, H. H., Wei, K. K., & Benbasat, I. (2003). Predicting intention to adopt interorganizational linkages: An institutional perspective. MIS Quarterly, 27(1), 19-49.

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. MIS Quarterly, 15(1), 125-142.

Thorn, B. K., & Connolly, T. (1987). Discretionary databases - A theory and some experimental findings. Communication Research, 14(5), 512-528.

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.

Tolbert, P. S. (1985). Institutional Environments and Resource Dependence: Sources of Administrative Structure in Institutions of Higher Education. Administrative Science Quarterly, 30(1), 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(1), 22-39.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27(3), 425-478.

Vogeli, C., Yucel, R., Bendavid, E., Jones, L. M., Anderson, M. S., Louis, K. S., & Campbell, E. G. (2006). Data withholding and the next generation of scientists: Results of a national survey. Academic Medicine, 81(2), 128-136.

Wasko, M., & Faraj, S. 2000. It is what one does: Why people participate and help others in electronic communities of practice. Journal of Strategic Information Systems, 9(2-3), 155-173.

Watson, S., & Hewett, K. (2006). A Multi-Theoretical Model of Knowledge Transfer in Organizations: Determinants of Knowledge Contribution and Knowledge Reuse*. Journal of Management Studies, 43(2), 141-173.

Weber, Nicholas M., Piwowar, Heather A., & Vision, Todd J. (2010). Evaluating data citation and sharing policies in the environmental sciences. Proceedings of the American Society for Information Science and Technology, 47(1), 1-2.

Witt, M. (2008). Institutional Repositories and Research Data Curation in a Distributed Environment. Library Trends, 57(2), 191-201.

Wu, W.-L., Lin, C.-H., Hsu, B.-F., & Yeh, R.-S. (2009). Interpersonal trust and knowledge sharing: Moderating effects of individual altruism and a social interaction environment. Social Behavior and Personality, 37(1), 83-93.

Ziman, J. (2000). Real Science: What It Is, and What It Means. Cambridge: Cambridge UP.

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(1-2), 5-16.

Zsidisin, G. A., Melnyk, S. A., & Ragatz, G. L. (2005). An institutional theory perspective of business continuity planning for purchasing and supply management. International Journal of Production Research, 43(16), 3401-3420.

Appendix A. Measurement Items for Research Constructs

| Construct | Items | Sources |
|---|---|---|
| Regulative Pressure by Funding Agencies | • Data sharing is mandated by the policy of public funding agencies.<br>• Data sharing policy of public funding agencies is enforced.<br>• Public funding agencies require researchers to share data.<br>• Public funding agencies can penalize researchers if they do not share data. | (Kostova et al., 2002)<br>(Teo et al., 2003) |
| Regulative Pressure by Journals | • Data sharing is mandated by journals' policy.<br>• Data sharing policy of journals is enforced.<br>• Journals require researchers to share data.<br>• Journals can penalize researchers if they do not share data. | (Kostova et al., 2002)<br>(Teo et al., 2003) |
| Normative Pressure | • It is expected that researchers would share data.<br>• Researchers care a great deal about data sharing.<br>• Researchers share data even if not required by policies.<br>• Many researchers are currently participating in data sharing. | (Kostova et al., 2002)<br>(Son et al., 2007) |
| Data Repository | • Researchers can easily access data repositories.<br>• Data repositories are available for researchers to share data.<br>• Researchers have the data repositories necessary to share data. | (Taylor et al., 1995)<br>(Venkatesh et al., 2003) |
| Perceived Career Benefit | • I can earn academic credit such as more citations by sharing data.<br>• Data sharing would enhance my academic recognition.<br>• Data sharing would improve my status in a research community.<br>• Data sharing would be helpful in my academic career. | (Wasko et al., 2000)<br>(Bock et al., 2005) |
| Perceived Career Risk | • There is a high probability of losing publication opportunities if I share data.<br>• Data sharing may cause my research ideas to be stolen by other researchers.<br>• My shared data may be misused or misinterpreted by other researchers.<br>• I believe that the overall riskiness of data sharing is high. | (Featherman et al., 2003)<br>(Pavlou 2003) |
| Perceived Effort | • Sharing data involves too much time for me (e.g. to organize/annotate).<br>• I need to make a significant effort to share data.<br>• I would find data sharing difficult to do.<br>• Overall, data sharing requires a significant amount of time and effort. | (Davis 1989)<br>(Davis et al., 1989)<br>(Thompson et al., 1991) |

| | | |
|---|---|---|
| Scholarly Altruism | • I am willing to help other researchers by sharing data.<br>• I would share data so that other researchers can conduct their research more easily.<br>• I would share data so that other researchers can utilize it for their research.<br>• I would share data to support open scientific research.<br>• I would share data to contribute to better scientific research.<br>• I would share data to help improve the quality of scientific research. | (Kankanhalli et al., 2005)<br>(Baytiyeh et al., 2010)<br>Newly Developed |
| Data Sharing Behavior | • How frequently have you deposited your data into disciplinary data repositories for every article?<br>• How frequently have you deposited your data into institutional data repositories for every article?<br>• How frequently have you uploaded your data into public Web spaces for every article?<br>• How frequently have you provided access to your data by publishing supplement materials for every article?<br>• How frequently have you responded to the data sharing request(s) by providing data via personal communication methods (e.g. email)? | Newly Developed |

Appendix B. Disciplines of Survey Participants

| Main Discipline | Sub Discipline | Frequency | Percentage |
|---|---|---|---|
| Engineering | Biomedical Engineering | 28 | 2.13% |
| | Chemical Engineering | 35 | 2.66% |
| | Civil Engineering | 27 | 2.05% |
| | Electrical Engineering | 26 | 1.97% |
| | Environmental Engineering | 22 | 1.67% |
| | Mechanical Engineering | 23 | 1.75% |
| | Metallurgical and Materials Engineering | 20 | 1.52% |
| Physical Sciences | Astronomy | 27 | 2.05% |
| | Chemistry | 30 | 2.28% |
| | Physics | 36 | 2.73% |
| Earth, Atmospheric, and Ocean Sciences | Atmospheric Sciences | 20 | 1.52% |
| | Geosciences | 52 | 3.95% |
| | Ocean Sciences | 42 | 3.19% |
| Agricultural Sciences | Agricultural Sciences | 26 | 1.97% |
| | Animal Sciences | 22 | 1.67% |
| | Forestry | 21 | 1.59% |
| | Natural Resources Conservation | 21 | 1.59% |
| | Plant Sciences | 39 | 2.96% |
| Biological Sciences | Biochemistry | 55 | 4.18% |
| | Biology | 21 | 1.59% |
| | Biometry and Epidemiology | 15 | 1.14% |
| | Biophysics | 24 | 1.82% |
| | Botany | 17 | 1.29% |
| | Cell Biology | 35 | 2.66% |
| | Developmental Biology | 32 | 2.43% |
| | Ecology | 60 | 4.56% |
| | Entomology and Parasitology | 21 | 1.59% |
| | Genetics | 48 | 3.64% |
| | Microbio, Immunology, and Virology | 70 | 5.32% |
| | Molecular Biology | 57 | 4.33% |
| | Neuroscience | 73 | 5.54% |
| | Physiology | 24 | 1.82% |
| Psychology | Clinical Psychology | 22 | 1.67% |
| | Psychology, Except Clinical | 34 | 2.58% |
| | Psychology, Combined | 21 | 1.59% |
| Social Sciences | Anthropology | 23 | 1.75% |
| | Geography | 23 | 1.75% |
| | Political Science | 30 | 2.28% |
| | Public Administration | 15 | 1.14% |
| | Sociology | 24 | 1.82% |
| Health Fields | Nursing | 21 | 1.59% |
| | Oncology/Cancer Research | 16 | 1.21% |
| | Preventive Medicine & Comm. Health | 19 | 1.44% |
| Total | | 1317 | 100% |

Appendix C. Survey Instrument

**(Page 1)**

**NOTE**: In this survey, ***Data Sharing*** means providing the raw data of your published articles to other researchers outside your research group(s) by making it accessible through data repositories/ public web spaces/ supplementary materials or by sending the data via personal communication methods upon request.

**ABOUT YOUR DISCIPLINE**

1. Which one of the following best describes your primary subject discipline based on your current research? *(Dropdown Selection Provided)*

Please indicate to what extent you agree with the following statements. For validation reasons, we may have to ask similar questions.

| **2. Public Funding Agencies** In my discipline, | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| Data sharing is mandated by the policy of public funding agencies. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing policy of public funding agencies is enforced. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Public funding agencies require researchers to share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Public funding agencies can penalize researchers if they do not share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| **3. Journal Publishers** In my discipline, | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| Data sharing is mandated by journals' policy. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing policy of journals is enforced. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Journals require researchers to share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Journals can penalize researchers if they do not share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| **4. Atmosphere** In my discipline, | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| It is expected that researchers would share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Researchers care a great deal about data sharing. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Researchers share data even if not required by policies. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Many researchers are currently participating in data sharing. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| **5. Data Repositories** | Strongly Agree |
|---|---|

| In my discipline, | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| Researchers can easily access data repositories. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data repositories are available for researchers to share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Researchers have the data repositories necessary to share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

**(Page 2)**

## ABOUT YOUR MOTIVATION

| **6. For Other Researchers** | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| I am willing to help other researchers by sharing data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I would share data so that other researchers can conduct their research more easily. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I would share data so that other researchers can utilize it for their research. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| **7. Benefits** | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| I can earn academic credit such as more citations by sharing data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing would enhance my academic recognition. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing would improve my status in a research community. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing would be helpful in my academic career. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| **8. Concerns** | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| There is a high probability of losing publication opportunities if I share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Data sharing may cause my research ideas to be stolen by other researchers. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| My shared data may be misused or misinterpreted by other researchers. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I believe that the overall riskiness of data sharing is high. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

**9. Efforts**

|  | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| Sharing data involves too much time for me (e.g. to organize/annotate). | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I need to make a significant effort to share data. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I would find data sharing difficult to do. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Overall, data sharing requires a significant amount of time and effort. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

**(Page 3)**

**ABOUT YOUR DATA SHARING BEHAVIOR**

**10. For Research Community**

|  | Strongly Disagree | Moderately Disagree | Slightly Disagree | Neutral | Slightly Agree | Moderately Agree | Strongly Agree | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| I would share data to support open scientific research. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I would share data to contribute to better scientific research. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| I would share data to help improve the quality of scientific research. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

11. In the last two years, how many publications involving actual research data have you produced per year?
a) None b) 1-2 c) 3-4 d) 5-6 e) 7+

**12. Data Sharing Frequencies**

In the last two years, how frequently have you…

|  | Never | Rarely | Occasionally | Sometimes | Frequently | Usually | Every time | Not Applicable | Do Not Know |
|---|---|---|---|---|---|---|---|---|---|
| Deposited your data into disciplinary data repositories for every article? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Deposited your data into institutional data repositories for every article? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Uploaded your data into “public” Web spaces for every article? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Provided access to your data by publishing supplement materials for every article? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Been personally asked to share data for each article? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Responded to the request(s) by providing data via personal communication methods (e.g. email)? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

**(Page 4)**

**ABOUT YOURSELF**

13. What is your age?
a) Under 24 b) 25-34 c) 35-44
d) 45-54 e) 55-64 f) 65+

14. What is your gender?
a) Male b) Female

15. What is your ethnic background?
a) Asian/Pacific Islander b) Black/African-American c) Caucasian
d) Hispanic e) Native American f) Other/Multi-Racial

16. What is your highest education so far?
a) Associate Degree b) Bachelor's Degree c) Master's Degree d) PhD/Doctoral Degree

17. What is your current position?
a) Assistant Professor b) Associate Professor c) Full Professor d) Professor Emeritus
e) Professor of Practice f) Lecturer/Instructor g) Post-Doctoral Fellow h) Researcher
i) Graduate Student j) Other (Specify)

18. Please choose the option most applicable to you.
a) Tenured b) On Tenure Track But Not Tenured c) Not on Tenure Track d) Retired

19. Which one of the following best describes your primary work sector?
a) Academic b) Government c) Commercial d) Non-Profit e) Other (Specify)

20. Please share any additional comments, questions, or suggestions about scientific data sharing.