Chapter 1

Predictors of Data Sharing and Reuse Behavior in Academic Communities

Jeonghyun Kim[†], Eric R. Schuler, and Anna Pechenina

University of North Texas

Denton, TX 76203 USA

[†]*E-mail: Jeonghyun.kim@unt.edu*

Research data sharing is one of the most interesting and challenging issues for researchers in the academic community. To investigate the effect of individual characteristics and organizational contexts on data sharing and reuse behaviors, this study employed a second analysis of the survey data. This study found that older researchers and those who allocate a lower percentage of their worktime to research are likely to share data and show a positive attitude toward data sharing. The study also found that academic researchers are likely to share data if their funding agency requires them to provide a data management plan and if their organization or project provides the necessary funds and processes to support data management.

Keywords: Data Sharing; Data Reuse; Academic Researcher

1. Introduction

The importance of sharing and reusing research data in academic communities has been well acknowledged. The value of data increases as they are aggregated into collections and as they become more available for reuse to address new research questions. Thus, shared data have served as a benchmark that allows others to study and refine methods of analysis and once collected, they can be creatively repurposed indefinitely by many hands and in many ways (Vision, 2010). As such, the ultimate goal of data sharing should be to increase scientific knowledge, but the benefits of sharing data will be achieved only if the available data are used or reused by others (Borgman, 2015a).

A new culture of data sharing has contributed to the significant crossroads that the academic community now is experiencing (Fenichel & Skelly, 2015). But sharing research data is still a complex issue presenting many challenges. There are both technical and social challenges that need to be addressed in any examination of research data access and sharing regimes (Arzberger *et al.*,

2004). Recognizing this challenge, there have been many initiatives driven by scientific journals, funding agencies, and policymakers to facilitate, fund, and request researchers to share their data. Further, many studies have been conducted to solve the challenges of data sharing, which include developing, applying, and adopting information and communication technologies, such as cyberinfrastructure, repository, metadata, and various tools that enable broad access to and optimal exploitation of research data (e.g., Conner, Ames, & Gill, 2013; Juve *et al.*, 2010; King, 2007; Poldrack & Gorgolewski, 2014); such studies focus on addressing the technical and operational aspects of data sharing. Still, little research devotes itself to the issue of data sharing in a comprehensive manner. Previous survey studies for understanding data-sharing patterns and practices provide insight into academic community members' data-sharing behavior (e.g., Fecher, Friesike, & Hebing, 2015; Tenopir *et al.*, 2010; Rathi, 2012; Wallis, Rolando, & Borgman, 2013), but do not facilitate direct measurement of data sharing and reuse behavior or its correlates.

The purpose of this study is to use a public-use data set from a survey conducted by the Usability and Assessment Working Group of the Data Observation Network for Earth (DataONE) project (Tenopir *et al.*, 2015a) to determine the individual characteristics and organizational contexts that affects the academic community's data sharing and reuse behavior. Specifically, the following research questions were addressed:

- (1) What are some individual characteristics and organizational contexts that affect academic data-sharing practice and attitude?
- (2) To what extent do individual characteristics and organizational contexts predict academic data-sharing practice and attitude?

2. Literature review

In light of the growing movement toward scientific data sharing, understanding which factors influence the use, reuse, and sharing of data by scientists is increasingly important. The literature on research data sharing behaviors has expanded rapidly in the past ten years. Several studies, both at the international and institutional levels, investigated how researchers and scientists in different disciplines use and share research data, as well as the socio-technical barriers preventing data sharing and how can they be overcome (Borgman, 2015b).

In recent years, several large-scale surveys have been conducted to investigate the patterns of data sharing around the world, such as PARSE.Insight (2009), a two-year project co-funded by the European Commission under the Seventh Framework Programme (FP7) Research

Infrastructures. The project conducted a survey to find out researchers' levels of data sharing, motivation for data sharing, and the barriers to data sharing they encountered. While the respondents to the PARSE.Insight surveys agreed that "data from publicly funded research should be shared and preserved," few were eager to share their research data with others. Furthermore, respondents expressed some reservations about the presence of barriers; for instance, some researchers made their data available for researchers in the same discipline, while some shared their data within their research collaborations and groups.

Similar to PARSE.Insight, Tenopir and her colleagues (2011) found that scientists were more willing to use other scientists' datasets rather than share their data. This finding highlighted a major gap between scientists' desire to use shared data and reluctance to share their own data. Most of the survey respondents (92%) agreed that their data should be cited and they should receive a proper citation credit. The survey also found that there was a significant difference in the opinions toward and practices of data sharing among scientists based on academic discipline and respondent's age. Tenopir and her colleagues (2015) conducted the follow-up study, where they identified some significant changes in views of data sharing and reuse by scientists. Overall, favorable perceptions of and practices surrounding data sharing has increased as the concerns over the risks associated with data sharing and reuse have arisen.

Kim and Stanton (2016) conducted a national survey of 1,317 scientists in 43 disciplines with the goal of evaluating the framework that combines institutional and individual factors that influence scientists' data-sharing behaviors. They employed planned behavior theory and institutional theory in the development of the test model, which includes two levels of predictors – institutional and individual. Institutional predictors included factors such as regulative pressures by funding agencies and journals, normative pressure, and availability of data repository. Individual predictors included factors such as perceived career benefit, perceived career risk, perceived effort, scholarly altruism, and data-sharing behavior. Overall, Kim and Stanton (2016) found that a combination of institutional and individual factors, namely regulative pressure by journals, normative pressure at the discipline level, and perceived career benefit and scholarly altruism, had a significant positive influence on data-sharing behaviors, while perceived effort by a researcher had a significant negative influence.

Linek *et al.* (2017) extended the idea of interplay between institutional and individual factors in academic data sharing to include the researcher's personality type. They surveyed 1,564 researchers across disciplines, looking at the relationship among social barriers and personality traits of the respondents

and its impact on data sharing attitudes. The surveys were conducted in English and German, and in addition to attitudes toward data sharing questions, contained standardized questions on selected personality factors, namely the so-called “Big Five, Machiavellianism and Social Desirability.” Linek and her colleagues found that data sharing is dependent on a combination of age, gender, and personality. Based on the findings, they stated that in order to foster data sharing attitudes, policy-making should be tailored to researchers’ individual needs and apprehensions as related to their personality types, rather than global interventions. In their opinion, “academic systems should establish circumstances in which the researchers’ personal values and ethical standards are better aligned with the public benefits of data sharing” (21).

In addition to large national and international projects, several institution-wide studies looking at individual characteristics of researchers were carried out to identify how they use and share research data. Wallis, Rolando, and Borgman (2013) conducted one such study over a ten-year period at the Center for Embedded Network Sensing (CENS). They found that CENS researchers expressed willingness to share their data under some conditions, specifically if they have the first right to publish the results from the data, received proper attribution as the data creators, knew the data requesters, and were expected to share their data by funding agencies. Akers and Doty’s study (2012, 2013) is also interesting to note; they interviewed faculty members at Emory University. The researchers found that at that time, most faculty members at Emory University did not share their research data outside their research group and did not deposit their data to any repositories or databanks. But they confirmed disciplinary differences among faculty members in terms of research data management actions, attitudes, and interest in data support services.

While the majority of studies focus on studying scientists and faculty researchers, there are a few that focus on graduate students as future researchers. Carlson and Stowell-Bracke (2013) examined the perceptions and practices of data sharing by graduate students working for agronomy researchers at the Water Quality Field Station (WQFS) at Purdue University. Interviews with six graduate students found that the graduate students would share their research data under certain conditions; the majority of them expressed concern about the timing of the release of research data and the potential of others to misuse or misunderstand their data.

Interestingly, it should be noted that studies discussed so far found that while academic researchers were generally willing to have their dataset shared with other researchers, actual data sharing and reuse practice is still questionable. This is because many of the surveys focused what researchers say

that they think rather than what they actually do. It is interesting to note a recently conducted study that reports the impact of shared data. Thelwall and Kousha (2017) analyzed data publishing and reuse statistics available through the Dryad repository for two evolutionary biology journals, *Evolution* and *Heredity*. They found that on the technical level data, publishing mandates worked well (at least for the two journals evaluated) and all of the authors of the articles within the journals shared their data. But the authors added that shared datasets were downloaded slowly over time and did not lead to more citations for the articles associated with the published data.

3. Methodology

3.1. Samples

This study involves secondary analysis of existing data collected by the Usability and Assessment Working Group of the Data Observation Network for Earth (DataONE) project survey (Tenopir *et al.*, 2015b). The survey was conducted as a follow-up study to their 2009/2010 baseline study (Tenopir *et al.*, 2011), which was conducted to identify the current state of data sharing and reuse perceptions and practices among research scientists. This follow-up survey received a total of 1,015 responses from multinational samples of scientific researchers who are employed in various work sectors, including academic, government, and non-profit.

As the focus of this study is the academic community, data from academic work sectors were extracted; as a result, a total of 746 respondents were selected. However, individuals who had more than 25% missing data or failed to complete the survey ($n = 264$) were excluded from the analyses, leaving a sample of 482 respondents. A description of the respondents appears in Appendix A.

3.2. Measures

Appendix A presents the operationalization of the key dimensions, attributes, and questions used in this study.

After a review of the detailed data coding, further data recoding was completed. Age was recoded into five groups based on developmental stages (Bjorklund, 2015, p. 356). The five stages consist of: (1) emerging adulthood (i.e., 10–24 years); (2) early adulthood (i.e., 25–39 years); (3) middle adulthood (i.e., 40–64 years); (4) older adulthood (i.e., 65–74 years); and late adulthood (i.e., 75 years and older). The work roles for emerging adulthood

are described as mostly consisting of entry-level jobs. In addition, whereas early adulthood is characterized by an emphasis on selecting a career and career establishment, middle adulthood can be considered the peak years of career success (Bjorklund, 2015, p. 356). The position was re-organized so that administrator was coded as 1; assistant/associate/professor/emeritus categories were coded as 2; undergraduate and graduate students were coded as 3; lecturer was coded as 4; post-doctoral fellow and researcher were coded as 5; and librarian/data manager/consultant were all coded as 6. The responses for the subject discipline were recoded based on the taxonomy of subjects (The National Academies of Sciences, Engineering, and Medicine, July 2016). As such, fields under Life Sciences were coded as 1; subjects under Physical Sciences and Math were coded as 2; Engineering fields were coded as 3; subjects under Social/Behavioral Sciences were coded as 4; and Arts and Humanities were coded as 5. The percentages of time spent on research and time spent on data management were also coded based on grouped percentages and quartile splits.

Furthermore, responses that were coded as “Not Sure” were left as missing data in the initial dataset. Since these were true responses and not technically missing, the values of “Not Sure” were re-coded to the dataset and scored next to the Neutral point option in the Likert data so as not to affect the interpretations of the scores. (For example, having “Not Sure = 6” as the highest score compared to the actual responses in the dataset). These changes to the “Not Sure” as a mid-point were recoded for variable series: Q15, Q17, Q18, Q19, Q20, and Q21. In particular, item 2 in Q15 was of an opposite valence compared to items 1 and 3, and as such was reversed coded.

3.3. Data preparation and pre-analysis

To determine the number of factors to retain for variables observed in this study, Principal Axis Factoring (PAF), which is based on theory and content of the questions, was conducted. It should be noted that this study employed PAF analyses since theory was utilized and the study was not trying to reduce the number of items into the smallest number of components (Henson & Roberts, 2006). For this analysis, missing values in the dataset were imputed using multiple imputation, which generated five multiply imputed datasets.

3.3.1. Derivation of the outcome variables

Outcome variables selected in this study are data-sharing practice, attitude toward data sharing and attitude toward data reuse. Multiple items were

employed to measure those outcome variables and items are scored using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Q15, which reflects data-sharing practice, consists of three items; a PAF with one factor retained was run. As a result, items 1 and 3 had high standardized pattern/structure coefficients, but item 2 was given a lower amount of credit in the creation of the factor score. Because of the low internal reliability and low pattern/structure coefficient, item 2 was removed. A PAF of items 1 and 3 was conducted with the one factor explaining 78.93% of the variance; this factor was labeled as “data sharing practice.”

A PAF was run with eight items for Q21, which includes five items for the respondents’ attitude toward data sharing (i.e., items 2, 3, 4, 6, and 7), and attitude toward data reuse (i.e., items 1, 5, and 8). The items with low pattern/structure coefficients were systematically removed from the analyses, leaving items 2, 3, and 6 in the final PAF. After the one-factor PAF, each of the three pattern/structure coefficients had a high correlation to the factor score. As such, the factor scores from the three items were retained for subsequent analyses and this factor was labeled as “attitude toward data sharing;” this factor explained 68.55% of the total variance.

Q20 consists of five items that reflect respondents’ attitude toward data reuse; two items indicate the respondents’ concern about lack of access to data (items 1 and 2) and three items reflect concern about misuse and misinterpretation of data (items 3, 4, and 5). Based on the content, there were two proposed factors that should be retained. The two factor PAF for Q20 with a varimax rotation was run; there was a simple structure, with items having high coefficients on one factor and near zero coefficients on the other factor. The factor scores for both factors were saved in the dataset for subsequent analyses, and they were labeled as “attitude toward data access (items 1 and 2)” and “attitude toward data misuse (items 3, 4, and 5).”

Table 1. Data sharing and reuse behavior factors

	Pattern structure coefficient	M	SD
Factor: Data-sharing practice			
I share my data with others	.76	4.01	1.06
Others can access my data easily	.76	3.00	1.36

Factor: Attitude toward data sharing

I would be willing to place at least some of my data into a central data repository with no restrictions	.81	4.20	1.03
I would be willing to place all of my data into a central data repository	.70	3.15	1.39
I would be willing to share data across a broad group of researchers	.68	4.30	.81

Factor: Attitude toward data access

Lack of access to data generated by other researchers or institutions is a major impediment to progress in science	.83	3.92	1.03
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions	.83	3.28	1.24

Factor: Attitude toward data re-use

Data may be misinterpreted due to complexity of the data	.79	4.07	.96
Data may be misinterpreted due to poor quality of the data	.81	4.13	.93
Data may be used in other ways than intended	.57	4.13	.90

3.3.2. Derivation of the predictor variables

For individual-level predictors, characteristics of demographic profiles, including age and position; disciplinary factors, including subject domain and types of data source used; and worktime spent on various activities are selected. For organizational-level predictors, funding agency related items are selected as it was assumed that pressure from funding agencies would influence data sharing; thus, respondents’ primary funding agency and whether their funding agency requires data management plans are selected. Additionally, their institutional support for data management in various areas was considered as institutional support for data management and data curation has a critical impact on scholar’s data-sharing behaviors (e.g., Jeng, He, & Oh, 2016). In the original study conducted by Tenopir and her colleagues, respondents were asked to what degree they agree or disagree with a series of statements.

A PAF was performed on Q17, Q18, Q19-1, and Q 19-2, which reflect how much the respondents’ organization or project is involved with data management, respectively, as those items were measured using multiple items. A single factor was extracted for each. As presented in Table 4, items for each question had high standardized pattern/structure coefficients and they were given a large amount of credit in creating the factor scores. The factor scores for each question were retained in the multiply imputed datasets. They were labeled as “institutional support for data management process,” “institutional support for data management training,” “institutional support for data management funding,” and “institutional support for data management infrastructure.”

Table 2. Institutional support factors

	Pattern structure coefficient	M	SD
Factor: Institutional support for data management process			
My organization has a formal process for managing data during the life of the project (short-term)	.88	2.80	1.41
My organization has a formal process for storing data beyond the life of the project (long-term)	.88	2.82	1.42
Factor: Institutional support for data management training			
My organization or project provides training on best practices for data management	.86	2.58	1.34
My organization or project provides assistance on creating data management plans	.88	2.59	1.37
My organization or project provides assistance on creating metadata to describe my data or datasets	.91	2.44	1.33
My organization or project provides training on how to cite datasets	.79	2.54	1.32
Factor: Institutional support for data management funding			
My organization or project provides the funds to support data management during the life of a research project (short-term)	.83	2.74	1.43

My organization or project provides the funds to support data management during the life of a research project (long-term)	.83	2.40	1.33
Factor: Institutional support for data management infrastructure			
My organization or project provides the software tools and technical support for data management during the life of the project (short-term)	.86	3.04	1.40
My organization or project provides the software tools and technical support for data management during the life of the project (long-term)	.86	2.70	1.36

3.4. Data analysis

To determine which of the independent variables could be used to predict academic researchers’ data sharing and reuse practice and attitude, a series of hierarchical multiple linear regression analyses were used. Along with individual characteristics (age, position, subject discipline, research worktime, data management worktime), organizational contexts such as primary funding agency, their mandates variables, and factor scores obtained for institutional support for data management were used as predictor variables. The factor scores for the following variables, which were constructed as a result of a factor analysis, were used as outcome variables: data-sharing practice, attitude toward data sharing, opinion on data access, and opinion on data misuse.

It also should be noted that a set of assumptions for regression analysis were reviewed. To assess for the linearity of the continuous independent variables and dependent variable factor scores, scatterplots were examined across each of the multiply imputed datasets. Additionally, the pooled correlational matrix with the additional continuous independent variables was examined. Many of the independent variables and dependent variables scatterplots appeared to have no relationship among the variables. These results were additionally supported by the pooled correlations for multiple regression.

The data analysis method chosen for the ultimate results of this study was a hierarchical multiple linear regression, which is useful when describing a relationship among two or more variables and when analyzing these variables according to a specific order. We entered predictor variables in two steps; in step 1, individual characteristics were controlled for when appropriate, then organizational context variables were entered in step 2.

See Appendix B for descriptive data for respondents’ individual characteristics and organizational contexts.

4. Results

4.1. Predictors of data sharing practice

As presented in Table 3, the first block of the hierarchical multiple linear regression looks at the significance of the individual-level predictors on data-sharing practice; it was statistically significant, ranging from $F(14, 457) = 4.20$, $p < .001$, to $F(14, 457) = 4.49$, $p < .001$. Four of the variables, including age, research worktime, data management worktime, and social science survey data, were statistically significant predictors of data-sharing practice.

In the second block of the hierarchical multiple linear regression, organizational context factors were added to the analysis. The organizational context factor block had a statistically significant improvement in accounting for variability across the imputed datasets, from $\Delta F(6, 451) = 3.52$, $\Delta p < .01$ to $\Delta F(6, 451) = 3.91$, $\Delta p < .01$. After including the organizational context factors, data management worktime was not significant, whereas research worktime became statistically significant. Age and social science survey data accounted for a significant amount of variance in the data-sharing practice factor score when holding all other variables constant. Whether the funding agency requires a data management plan was significant, as was institutional support for data management funding.

Table 3. Hierarchical multiple regression analysis for variables predicting data-sharing practice

Predictors	Step 1			Step 2		
	B	S.E.	t	B	S.E.	t
Age	.18	.06	3.13**	.17	.06	3.12**
Position	.01	.03	.23	.01	.03	.45
Discipline	-.05	.03	-1.62	-.05	.03	-1.56
Research Worktime	-.07	.04	-2.02*	-.07	.04	-2.03*
Data Management (DM) Worktime	.07	.04	2.12*	.05	.04	1.50
Data-Abiotic Surveys	.03	.10	.30	-.00	.10	-.02
Data-Biotic Surveys	.12	.10	1.56	.11	.10	1.13
Data-Data Models	.09	.08	1.07	.09	.08	1.11
Data-Experimental	-.03	.08	-.38	-.02	.08	-.23

Data-Interviews	-.01	.12	-.08	.05	.12	.39
Data-Observational	.06	.08	.73	.04	.08	.50
Data-Remote-sensed abiotic data	.14	.10	1.36	.13	.10	1.28
Data-Remote-sensed biotic data	.17	.11	1.52	.14	.11	1.28
Data-Social Science Surveys	-.23	.11	-2.07*	-.27	.11	-2.48*
Primary Funding Agency				-.03	.02	-1.64
Funding Agency's Mandates				.12	.06	2.13*
Institutional Support for DM Process				-.01	.05	-.22
Institutional Support for DM Training				.01	.05	.18
Institutional Support for DM Funding				.15	.06	2.64**
Institutional Support for DM Tools/Technology				-.02	.06	-.03

* $p < .050$, ** $p < .010$

4.2. Predictors of attitude towards data sharing

Table 4 presents the results of a hierarchical multiple linear regression, which was utilized to assess how individual characteristic factors and organizational context factors could account for variability in attitude toward data sharing and reuse. The first block of individual factors in predicting data attitude was statistically significant, ranging from $F(14, 457) = 2.30, p < .01$ to $F(14, 457) = 2.58, p < .05$. Only the individual factor of research worktime was statistically significant; as such, the more time spent on research, the less positive attitude toward data sharing.

The institutional predictor block (Block 2) had a statistically significant improvement in accounting for variability across the five imputed datasets, from $\Delta F(6, 451) = 2.22, \Delta p < .05$ to $\Delta F(6, 451) = 2.59, \Delta p < .05$; including the institutional factors improved the variance in attitude factor scores by approximately 5%. After including the institutional context factors, research worktime was still statistically significant. Of the institutional factors, whether the funding agency requires a data management plan and if the institution provided training on best practices for data management were significant predictors.

Table 4. Hierarchical multiple regression analysis for variables predicting attitude toward data sharing

Predictors	Block 1			Block 2		
	B	S.E.	t	B	S.E.	t
Age	.03	.06	.45	.01	.06	.09
Position	.03	.03	1.03	.05	.03	1.63
Discipline	-.02	.03	-.54	-.02	.03	-.61
Research Worktime	-.08	.04	-2.01*	-.08	.04	-2.23*
Data Management Worktime	.03	.04	.93	.03	.04	.68
Data-Abiotic surveys	.15	.11	1.43	.13	.11	1.22
Data-Biotic surveys	.09	.11	.86	.08	.11	.99
Data-Data models	.09	.09	1.05	.09	.09	1.07
Data-Experimental	-.06	.08	-.67	-.06	.08	-.76
Data-Interviews	-.18	.12	-1.47	-.18	.12	-1.48
Data-Observational	.07	.09	.85	.06	.09	.69
Data-Remote-sensed abiotic data	.05	.11	.50	.01	.11	.04
Data-Remote-sensed biotic data	.10	.12	.88	.13	.12	1.09
Data-Social Science Surveys	-.02	.11	-.21	-.03	.11	-.29
Primary Funding Agency				-.02	.02	-1.19
Funding Agency's Mandates				.15	.06	2.51**
Institutional Support for DM Process				-.12	.06	-2.12*
Institutional Support for DM Training				.02	.06	.36
Institutional Support for DM Funding				.00	.06	.06
Institutional Support for DM Infrastructure				.03	.06	.47

* $p < .050$, ** $p < .010$

4.3. Predictors of attitude toward data reuse

A hierarchical multiple linear regression was utilized to assess how individual and organizational predictors could account for variability in attitude toward data reuse.

The first block of individual factors in predicting attitude toward data misuse was statistically significant, ranging from $F(14, 457) = 2.46, p < .01$ to $F(14, 457) = 2.57, p < .001$. Within the individual factors, there were two predictors that were statistically significant: remote-sensed abiotic surveys were used and age based on psychological development groups. The institutional predictor block (Block 2) did not have a statistically significant improvement in accounting for variability across the five imputed datasets, from $\Delta F(6, 451) = .35, \Delta p = .91$ to $\Delta F(6, 451) = .40, \Delta p = .88$. As such, the second block of predictors were not interpreted. A list of all the pooled coefficients of the individual factors are in Table 5.

Table 5. Hierarchical multiple regression analysis for variables predicting attitude towards data misuse

Predictors	Block 1		
	B	S.E.	t
Age	-.19	.07	-2.92**
Position	.00	.03	.42
Discipline	-.03	.04	-.78
Research Worktime	-.70	.04	-1.60
Data Management Worktime	-.06	.04	1.45
Data-Abiotic surveys	.01	.12	.08
Data-Biotic surveys	-.04	.12	-.33
Data-Data models	.08	.09	.81
Data-Experimental	-.05	.09	-.57
Data-Interviews	-.08	.13	-.56
Data-Observational	.03	.09	.28
Data-Remote-sensed abiotic data	.26	.12	2.16*

Data-Remote-sensed biotic data	.03	.13	.20
Types of Data-Social Science Surveys	.02	.13	.20

* $p < .050$, ** $p < .010$

5. Discussion

At the individual level, age and research worktime were found to have a significant, positive effect on academic scientists’ data sharing and reuse behaviors. Older people are more likely to make their research data readily accessible to others, and they agreed that their own data are easy to access. In other words, younger people have a more conservative attitude toward data sharing and reuse; they are more concerned about the complexities and circumstantial matters in the creation and use of data. Those who spend more time in research and are expected to produce more research data tend to have a more restrained attitude toward data sharing as well.

Although many previous studies confirmed that disciplinary data practices, norms, and policies vary significantly and some disciplines publish data more readily than others (e.g., Borgman, 2007; Carlson & Anderson, 2007), we did not observe any clear disciplinary differences in data practices and behaviors. However, the data sources used by academic researchers were found to have a significant relationship with data sharing and reuse behaviors. Those who identified themselves as social science data users tend to be resistant to data sharing and reuse; this may be because social sciences emphasize the protection of study participants (Curty *et al.*, 2016). Further, individuals and small teams in social sciences collect data for specific research studies; these data tend to be small in volume, local in character, intended for use only by these teams, and are less likely to be structured in ways that allow data to be transferred easily between individuals or teams. Remotely-sensed abiotic data users are more concerned about potential data misuse and misinterpretation. Remotely-sensed data collected from satellite or airborne platforms, such as meteorological data, are an important basis for the country; the public welfare scientific data resources, not only play an important role in the weather, weather forecasting, climate change, and other meteorological service and scientific activities, but are important reference information for macro decision-making, national economic construction, scientific and technological innovation, and public life (Du, Xin, & Wu, 2013). For this reason, those remotely-sensed abiotic data users might be more considerate toward the use and reuse of data.

Among organizational context variables, this study confirms that data management requirements by the funding organizations influenced academic scientists' data-sharing behaviors positively. Those who are required to provide a data management plan that spells out what data will be generated and how those data will be managed and shared both during and after the duration of any grant, are likely prone to make their data open and available to others. This supports previous studies demonstrating that federal data-sharing policies and data management plan mandates have created peer pressure for researchers to share and curate data (McCullough, McGeary, & Harrison, 2008; Piwowar & Chapman, 2010).

Out of other institutional context factors, institutional support for data management funding and process was found to have a significant effect on academic researchers' data-sharing behaviors. Those who responded that their organization provides the funds to support data management are likely to share their data. In a similar vein, those who responded that their organization has a formal process for managing and storing data tend to have positive attitude toward data sharing. However, organizational support for data management assistance/training and support for data management infrastructure was not found to have a relationship with data-sharing behaviors. This may be the result of the overall low level of institutional support observed in this study; nearly half the respondents reported that their organization does not provide any training or assistance for data management. Also, more than 35% of the respondents replied that their organization and project do not provide tools and software for data management. It should be noted that more than 10% of the respondents did not know whether their organization offers such support services or not. This implies that there still is a gap between academic researchers and librarian perspectives of the availability of support services. Another possible interpretation of this insignificant result is that since the question items related to training and assistance for data sharing and publishing are not included specifically, the effect of institutional support was not fully reflected.

6. Conclusion

This secondary analysis of the data from the Usability and Assessment Working Group of the Data Observation Network for Earth (DataONE) project survey (Tenopir *et al.*, 2015b) suggests that both individual characteristics and organizational contexts need to be considered to encourage academic researchers' data-sharing and reuse behaviors. The most notable finding is that

institutional support for data management is one of keys to data sharing practice and attitude.

Further, this study calls for funding agencies’ continuous enforcement of data management and sharing policies as this study found that regulatory pressure exhibited by funding agencies does have a significant effect on academic researchers’ data-sharing practice and attitude. This implies that funding agencies may need to develop more systematic mechanisms or systems to encourage academic researchers’ data-sharing behaviors. Additionally, this study suggests academic institutions should consider developing their data policy to promote the prompt and open exchange of research data. Along with such a policy, institutions should develop and promote a robust and cohesive research data service, which provides their institution community with the expertise, tools, and infrastructure necessary to manage, steward, and share research data.

Appendix A. Dimensions, Attributes, and Questions

Dimensions	Attributes	Questions from the Original Survey
Individual-level Predictors	Age	Q6. Year of birth – Please select from the list below
	Position	Q4. My current position is:
	Subject Discipline	Q3. Which one of the following best describes your primary subject discipline?
	Research Worktime	Q2. What percentage of your work time is allocated to research activity?
	Data Management Worktime	Q2. What percentage of your work time is allocated to data management activity?
	Type of Data Source Used	Q7. Which terms best describe the type of data you use?
Organizational-level Predictors	Primary Funding Agency	Q9. Which of the following best describes the primary funding agency for your research?
	Funding Agency’s Mandates	Q10. Does your primary funding agency require you to provide a data management plan?

	Institutional Support for Data Management Process	<p>Q17. My organization has a formal process for.....</p> <ol style="list-style-type: none"> 1) managing data during the life of the project (short-term) 2) storing data beyond the life of the project (long-term)
	Institutional Support for Data Management Training	<p>Q18. My organization or project provides training or assistance on</p> <ol style="list-style-type: none"> 1) training on best practices for data management 2) assistance on creating data management plans 3) assistance on creating metadata to describe my data or datasets 4) training on how to cite datasets
	Institutional Support for Data Management Funding	<p>Q 19-1. My organization or project provides the funds to support data management</p> <ol style="list-style-type: none"> 1) during the life of a research project (short-term) 2) beyond the life of a research project (long-term)
	Institutional Support for Data Management Infrastructure	<p>Q 19-2 My organization or project provides the software tools and technical support for data management</p> <ol style="list-style-type: none"> 1) during the life of the project (short-term) 2) beyond the life of the project (long-term)
Data Sharing and Reuse Behavior	Data-Sharing Practice	<p>Q15. The following statements relate to data sharing and access. Tell us how much you agree with each statement.</p> <ol style="list-style-type: none"> 1) I share my data with others 2) Others need my permission to access my data 3) Others can access my data easily

Attitude Toward Data Sharing	<p>Q21. The following statements relate to sharing scientific data. Tell us how much you agree with each statement.</p> <ol style="list-style-type: none">1) I would use other researchers' datasets if their datasets were easily accessible2) I would be willing to place at least some of my data into a central data repository with no restrictions3) I would be willing to place all of my data into a central data repository with no restrictions4) I would be more likely to make my data available if I could place conditions on access5) I am satisfied with my ability to integrate data from disparate sources to address research questions6) I would be willing to share data across a broad group of researchers7) It is important that my data are cited when used by other researchers8) It is appropriate to create new datasets from shared data
Attitude Toward Data Reuse	<p>Q20. The following statements relate to your views on the use of scientific research data. Tell us how much you agree with each statement.</p> <ol style="list-style-type: none">1) Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.2) Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.3) Data may be misinterpreted due to complexity of the data.4) Data may be misinterpreted due to poor quality of the data.5) Data may be used in other ways than intended.

Appendix B. Descriptive Statistics

Variable	Frequency (Percentage)
Age	
10-24 years old	4 (.8%)
25-39 years old	177 (36.7%)
40-64 years old	245 (50.8%)
65-74 years old	35 (7.3%)
75 years and older	11 (2.3%)
Chose not to disclose	10 (2.1%)
Position	
Administrator	16 (3.3%)
Professor	226 (46.9%)
Student (Graduate/Undergraduate)	71 (14.7%)
Lecturer	14 (2.9%)
Research/Post-Doctoral Fellow	128 (26.6%)
Librarian/Data Manager	12 (2.5%)
Other	15 (3.1%)
Subject Domain	
Life Sciences	152 (31.5%)
Physical Sciences and Math	139 (28.8%)
Engineering	69 (14.3%)
Social/Behavioral Sciences	91 (18.9%)

Arts/Humanities	5 (1%)
Other	26 (5.4%)
Research Worktime	
0-25%	151 (31.3%)
25.1%-50%	152 (31.5%)
50.1% - 75%	61 (12.7%)
75.1% - 100%	118 (24.5%)
Data Management Worktime	
0-25%	185 (38.4%)
25.1%-50%	91 (18.9%)
50.1% - 75%	92 (19.1%)
75.1% - 100%	114 (23.6%)
Types of Data Source Use	
Abiotic surveys (soils, microclimate, hydrology, etc.)	164 (34%)
Biotic surveys	152 (31.5%)
Data models	202 (41.9%)
Experimental (involving some degree of manipulation)	238 (49.4%)
Interviews	103 (21.4%)
Observational (no manipulation involved)	255 (52.9%)
Remote-sensed abiotic data	174 (36.1%)
Remote-sensed abiotic data	115 (23/9%)
Social Science Survey	109 (22.6%)

Primary Funding Agency	
Federal/National Government	318 (66%)
State/Regional Government	44 (9.1%)
Local Government	3 (.6%)
Corporation	10 (2.1%)
Private Foundation	19 (3.9%)
Internal/My Institute	67 (13.9%)
Other	20 (4.1%)
Prefer not to disclose	1 (.2%)
Funding Agency Mandates	
Don't Know	75 (15.6%)
No	201 (41.7%)
Yes	191 (39.6%)
Prefer not to disclose	15 (3.1%)

References

Akers, K. G., & Doty, J. (2012). Differences among faculty ranks in view on research data management. *IASSIST Quarterly*, 36(2), 16-20.

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *The International Journal of Digital Curation*, 8(2), 5-26.

Arzberger, P. W., Schroeder, P., Beaulieu, A., Bowker, G. C., Casey, K., Laaksonen, L., & Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3(29), 135-152.

Bjorklund, B. R. (2015). *The journey of adulthood*. Upper Saddle River, NJ: Pearson Education.

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

- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078. doi:10.1002/asi.22634
- Borgman, C. L. (2015a). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: The MIT Press.
- Borgman, C. L. (2015b). If data sharing is the answer, what is the questions? *ERCIM News*, 100. Retrieved from <http://ercimnews.ercim.eu/en100/special/if-data-sharing-is-the-answer-what-is-the-question>
- Carlson, J., & Stowell-Bracke, M. (2013). Data management and sharing from the perspective of graduate students: An examination of the culture and practice at the Water Quality Field Station. *Portal: Libraries and the Academy*, 13(4), 343-361. doi:10.1351/pla.2013.0034
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12(2), 635-651.
- Conner, L. G., Ames, D. P., & Gill, R. A. (2013). HydroServer Lite as an open source solution for archiving and sharing environmental data for independent university labs. *Ecological Informatics*, 18, 171-177. <https://doi.org/10.1016/j.ecoinf.2013.08.006>
- Curty, R., Yoon, A., Jeng, W., & Qin, J. (2016). Untangling data sharing and reuse in social sciences. *Proceedings of the 79th Annual Meeting of the Association for Information Science and Technology*.
- Doolan, D. M., & Froelicher, E. S. (2009). Using an existing data set to answer new research questions: A methodological review. *Research and Theory for Nursing Practice*, 23(3), 203-215.
- Du, J., Xin, J., & Wu, S. (2013). Research on meteorological data sharing technology based on service oriented architecture. *Proceedings of the 2nd International Conference on Innovative Computing and Cloud Computing*.
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE*, 10(2), e0118053. doi:10.1371/journal.pone.0118053
- Fenichel, E. P., & Skelly, D. K. (2015). Why should data be free: Don't you get what you pay for? *BioScience*, 65(6), 541-542. <https://doi.org/10.1093/biosci/biv052>
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393-416. doi:10.1177/0013164405282485
- Jeng, W., He, D., & Oh, J. S. (2016). Toward a conceptual framework for data sharing practices in social sciences: A profile approach. *Proceedings of the 79th Annual Meeting of the Association for Information Science and Technology*.
- Juve, G., Deelman, E., Vahi, K., Mehta, G., Berriman, B., Herman, B. P., & Maechling, P. (2010). Data sharing options for scientific workflows on Amazon EC2. *SC'10 Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*.

- 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(4), 776-799.
- King, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research*, 36(2), 173-199. doi:10.1177/0049124107306660
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.
- Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLoS ONE*, 12(8): 0183216.
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics*, 41(4), 1406-1420.
- Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics*, 4(2), 148-156.
- Poldrack, R. A., & Gorgolewski, K. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, 17(11), 1510-1517, doi:10.1038/nn.3818
- Rathi, V., Dzara, K., Gross, C. P., Hrynaskiewicz, I., Joffe, S., Krumholz, H. M., & Kelly, M. S. (2012). Sharing of clinical trial data among trialists: A cross sectional survey. *BMJ*, 345, e7570. doi:10.1136/bmj.e7570
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., & Dorsett, K. (2015a). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, 10(8), e0134826. doi:10.1371/journal.pone.0134826
- 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), e21101. doi:10.1371/journal.pone.0021101
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., & Dorsett, K. (2015b). Data from: Changes in data sharing and data reuse practices and perceptions among scientists worldwide. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.1ph92>
- The National Academies of Science, Engineering, and Medicine (2006, July 31). *Taxonomy of fields and their subfields*. Retrieved from http://sites.nationalacademies.org/PGA/Resdoc/PGA_044522
- The PARSE.Insight (2009). *First insights into preservation of research output in Europe*. Retrieved from http://libereurope.eu/wp-content/uploads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf
- Thelwall, M. & Kousha, K. (2017). Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib Journal of Information Management*, 69(1), 36-45.

- Vision, T. J. (2010). Open data and the social contract of scientific publishing. *BioScience*, 60(5), 330-330.
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8(7), e67332. doi:10.1371/journal.pone.0067332