2013

Institutional and Individual Influences on Scientists' Data Sharing Behaviors

Youngseek Kim

Abstract

Institutional and Individual Influences on Scientists' Data Sharing Behaviors

by

Youngseek Kim

In modern research activities, scientific data sharing is essential, especially in terms of data-intensive science and scholarly communication. Scientific communities are making ongoing endeavors to promote scientific data sharing. Currently, however, data sharing is not always well-deployed throughout diverse science and engineering disciplines. Disciplinary traditions, organizational barriers, lack of technological infrastructure, and individual perceptions often contribute to limit scientists from sharing their data. Since scientists' data sharing practices are embedded in their respective disciplinary contexts, it is necessary to examine institutional influences as well as individual motivations on scientists' data sharing behaviors.

The objective of this research is to investigate the institutional and individual factors which influence scientists' data sharing behaviors in diverse scientific communities. Two theoretical perspectives, institutional theory and theory of planned behavior, are employed in developing a conceptual model, which shows the complementary nature of the institutional and individual factors influencing scientists' data sharing behaviors. Institutional theory can explain the context in which individual scientists are acting; whereas the theory of planned behavior can explain the underlying motivations behind scientists' data sharing behaviors in an institutional context.

This research uses a mixed-method approach by combining qualitative and quantitative methods: (1) interviews with the scientists in diverse scientific disciplines to understand the extent to which they share their data with other researchers and explore institutional and individual factors affecting their data sharing behaviors; and (2) survey research to examine to what extent those institutional and individual factors influence scientists' data sharing behaviors in diverse scientific disciplines.

The interview study with 25 scientists shows three groups of data sharing factors, including institutional influences (i.e. regulative pressures by funding agencies and journals and normative pressure); individual motivations (i.e. perceived benefit, risk, effort and scholarly altruism); and institutional resources (i.e. metadata and data repositories). The national survey (with 1,317 scientists in 43 disciplines) shows that regulative pressure by journals; normative pressure at a discipline level; and perceived career benefit and scholarly altruism at an individual level have significant positive relationships with data sharing behaviors; and that perceived effort has 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.

Institutional and Individual Influences on Scientists' Data Sharing Behaviors

Youngseek Kim

B.A., Seoul National University, 2006

M.S., Syracuse University, 2008

Dissertation

Submitted to the Graduate School of Syracuse University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Information Science and Technology

June 2013

Acknowledgements

First of all, I would like to thank all those who participated in this dissertation research, including 25 interviewees, 8 subject matter experts, 29 pretest participants, 34 pilot-test participants, and 2,470 national survey participants. This dissertation would not have been possible or completed without their support. In addition, I would like to acknowledge the CoS (Community of Scientists) Pivot for allowing me to use their scholar database in recruiting the survey participants.

I wish to express my sincere gratitude to my advisor, Dr. Ping Zhang, for her support and encouragement throughout my doctoral study and this dissertation work. She always provided a prompt response and constantly encouraged me to finish this dissertation. I would also like to offer special thanks to Dr. Jeffrey Stanton for his invaluable advice and guidance in this dissertation research. His research philosophy as well as insightful feedback significantly influenced this dissertation.

I am very grateful to my other committee members, Dr. Kevin Crowston, Dr. Jian Qin, Dr. Charles Driscoll, Dr. Paul Morarescu, and Dr. James Bellini, for their valuable comments and suggestions, which helped me improve the quality of my dissertation. I am also grateful to Dr. Murali Venkatesh for assisting me in understanding institutional theory, which is a critical part of the theoretical framework of this dissertation.

I would like to thank my doctoral colleagues, professors, and staff members in the iSchool at Syracuse University for their care and motivation during my graduate studies. I especially would like to thank Ms. Bridget Crary for making my doctoral study and dissertation process smoother. Additionally, I also would like to acknowledge Ms. Diane Stirling for proofreading and editing this dissertation.

Finally, I wish to express my warmest and utmost gratitude to my parents, HongJin Kim and BongSoo Aeo, for their priceless and infinite love. Their recognition of the importance of education enabled me to have academic curiosity and pursue my doctoral degree. Most importantly, I would like to recognize the invaluable support and patience of my wife, HyoKyung Kwak. Her encouragement and support throughout my dissertation work allowed me to finish this dissertation and achieve my doctoral degree.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Problem Statement

Research background, motivation of this research, definitions of terms, research objective and questions, theoretical perspective, and significance of the study are discussed in this chapter. The main objective of this research is to investigate the factors influencing scientists' data sharing behaviors in diverse scientific disciplines. In order to fully understand scientists' data sharing, I propose the research framework combining institutional and individual perspectives; it can explain how individual scientists make their decisions under institutional influences. The significance of this research is presented in terms of theory, method, and practice.

1.1. Background

Data sharing is a critical issue in modern scientific research with the emergence of e-Science or cyberinfrastructure. The term e-Science is defined as "networked and data-driven science," (Hey et al. 2006) and a critical aspect of it centers on global collaboration in key areas of science being enabled by grid computing and data-centric scientific research based on data repositories (Hey et al. 2002). e-Science promises to reshape and enhance the way science is done, by empowering data-driven scientific research and improving the synthesis and analysis of scientific data in a collaborative and shared fashion (Wright et al. 2011).

The underlying foundation of e-Science–data sharing–was enabled and facilitated by many contemporary scientific endeavors, including the development of networked collaboration technologies, institutional data repositories, and collaborative efforts on metadata standards. First, the advancement of networked collaboration technologies has enhanced the way scientists currently access information, communicate, and collaborate (Kling et al. 2000; McCain 2000). Second, the rise of institutional data repositories has helped scientists to share their data and novel scientific findings because they could better examine relationships among previous findings. Third, the collaborative efforts on data or metadata standards have increased accessibility of scientific data by different scientists. In summary, e-Science has revolutionized the process of scientific discovery by enabling data-centric science or scientists sharing their data and reusing others' data through technological development and collaborative effort (Hey et al. 2008).

The vision of data-intensive scientific research is made possible by sharing raw data sets among scientists. An enormous amount of primary data continues to be generated by large science institutions and individual scientists through new scientific research methods, such as simulations, sensor networks, and satellite surveys in different research fields (Hey et al. 2006). This huge amount of shared scientific data can potentially provide dramatic insights which cannot be found by looking only at individual data sets (Buetow 2005; Hey et al. 2006). Government agencies and research institutions promote data sharing through data repositories, where scientists can openly share their raw data (Atkins et al. 2003). Hey and Trefethen (2003) also highlight that the imminent availability of primary data sets through data repositories is one of the critical components which supports e-Science or cyberinfrastructure.

In the same vein, as science and engineering research becomes more data-intensive, data sharing and reuse appear to be important issues of scholarly communication in science and engineering fields (Cragin et al. 2006). Traditionally, scholarly knowledge was shared through journal articles or increasingly article pre-prints; however, diverse e-

Science technologies based on the Internet (e.g. personal communication methods and data repositories) allow scientists to share all their knowledge, especially raw data sets. In the perspective of scholarly communication, primary data collected by individual scientists becomes an important “information currency” along with research analyses and findings in the traditional publications (Davis et al. 2007). Individual scientists benefit from data sharing by validating previous research findings, developing new hypotheses, expediting their research works, and educating science trainees based on the shared raw data sets (Borgman 2007; Borgman 2010; Campbell et al. 2002; Fienberg 1994; Fienberg et al. 1985; Tenopir et al. 2011; Vickers 2006).

In order to achieve the core vision of data-intensive science, it is critical to allow individual scientists to share their research data with other scientists through diverse methods (i.e. data repositories or personal communications). Individual scientists usually work on small science or on their own research projects in a small or middle-sized group of graduate students, post-docs, and staff members. Individual scientists generate large amounts of data through their daily research activities (Boyce et al. 2006). Heidorn (2008) found that “(up to) 80% of all science is in the long tail of scientific research made up of smaller, less costly projects.” Carlson (2006) also argued that typically small science generates more data than big science, which requires high-cost resources and joint collaborations from multiple disciplines. Additionally, the scientists in small science span more scientific fields and generate increased and diverse forms of data over the researchers in big science (Carlson 2006).

Scientific data are more valuable when they are shared and can be reused beyond the value of when the data were originally collected (Faniel et al. 2011). In modern scientific research, it has become necessary for individual scientists to share their data with other scientists by using central or local repositories and/or personal communication methods. Within the last few decades, scientists observed the importance of data sharing, and many scientific communities paid considerable attention to the benefits of data sharing (Strier et al. 2010) because of the premise that data sharing would contribute to the advancement of science. Individual scientists' data sharing behaviors are more important in small science as compared to big science, which has systematic procedures of data management and institutional data repositories for data sharing. Small science often does not have any substantial mechanism and data repository to manage the growing amounts of data by individual scientists (Borgman et al. 2007b). This research focuses on scientists' data sharing behaviors in the context of small science rather than big science.

1.2. Motivation

As the raw data becomes important in terms of scholarly communication and data-intensive scientific research, data sharing is now essential in most modern research activities (Faniel et al. 2011). In terms of scholarly communication, the advancement of information and communication technologies has enabled scientists to share their data with their research publications for diverse purposes, including validating original research findings, building new hypotheses, expediting current research, and educating science trainees. Furthermore, in terms of data-intensive scientific research, data sharing can accelerate scientific collaboration and enable large-scale research. Borgman (2007) highlights how synthesized data for an initial research project can be raw data for subsequent research. Scientists can extend their research by conducting comparative studies and with more sophisticated analyses and syntheses that is based on shared data sets.

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). Also, many scientific journals' data sharing policies began to mandate data sharing for the published articles, which was implemented throughout several scientific communities (Faniel et al. 2011). Along with mandatory data sharing policies, scientific communities developed data repositories where scientists could freely and openly share their data, and also worked towards the development of metadata which facilitate data sharing.

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 et al. 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 which have good support and an environment towards 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 1968). 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 which facilitated scientists' data sharing. However, diverse external issues, including the policies developed by funding agencies, journals, and university tenure and promotion systems, 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.

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 which 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.

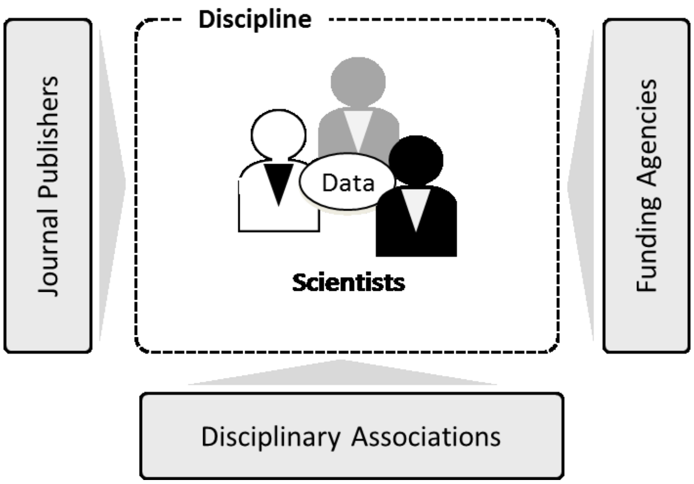


Figure 1.1 Scientists under Disciplinary Contexts

As seen in Figure 1.1 above, scientists' data sharing is embedded in their respective disciplinary contexts, including relevant associations, journal publishers, and funding agencies. For that reason, it is necessary to examine disciplinary influences on data sharing behaviors in diverse scientific disciplines. Scholars argue that data sharing is deeply rooted in the disciplinary practice and culture where scientists conduct their research (Sterling et al. 1990; Tenopir et al. 2011), and the facilitators or barriers vary significantly among and within scientific disciplines (Borgman 2007; Pryor 2009; Tenopir et al. 2011). Individual scientists' data sharing behaviors are influenced by institutional contexts which differ among disciplines. Both individual and institutional factors influencing scientists' data sharing need to be investigated carefully since this investigation can provide a holistic view of data sharing across diverse scientific disciplines.

Although the idea of data sharing is promising and can enhance scientific discovery, it cannot be achieved without scientists' voluntary data sharing behaviors and institutional supports. In order to achieve the core vision of data-intensive scientific research, it is critical to deploy data sharing among scientists. Successful data sharing can be achieved by considering technological infrastructure, institutional context, and individual motivations, which vary across disciplines (Borgman 2007; Pryor 2009; Tenopir et al. 2011). This study helps to understand the main factors which influence scientists' data sharing behaviors across different disciplines by considering both individual motivation and institutional contexts (including technological infrastructure) together.

1.3. Definitions of Terms

Small Science versus Big Science

Small science refers to science performed by an individual scientist or a small group of scientists (e.g. an investigator with a mix of post-docs, graduate students, and/or staffs) working on their own chosen projects. By contrast, big science refers to science performed by a significant number of scientists requiring huge amounts of resources and addressing large-scale scientific problems. In big science, scientists' decisions on data sharing are significantly restricted by the organizational policies of higher level decision makers. But in small science, scientists' decisions on data sharing are made by the individual scientists. This research examines individual scientists' decision making toward data sharing in their daily scientific research activities, so small science is a main context for this research.

Scientist

Scientist refers to a scholar or researcher in academia who generates and disseminates scientific knowledge publicly. STEM (Science, Technology, Engineering, and Mathematics) researchers are considered as the main group of scientists.

Research Data

Research data (data in general) refers to the extensive range of research results and relevant information. In the perspective of small science, individual scientists or a small group of scientists, collect data by using diverse collection methods including observation, experiment, and simulation. The research data may include any research-related information, such as research techniques and related materials (Blumenthal et al. 2006). Data are considered to be a fundamental infrastructural component of scientific research (Uhlir 2010), especially because in the perspective of data-intensive research, data are not the end products of research, but needs to be considered as part of an evolving data stream in a scientific field (Hilgartner et al. 1994).

Data Sharing

Data sharing is individual scientist's behavior to provide their raw (or preprocessed) data to other scientists by making it accessible through central/local data repositories or by sending data via personal communication methods upon request. In this research, data sharing does not involve providing data by big science research centers, which sometimes collect and distribute data to other scientists in their fields as their main duties.

Data Reuse

Data reuse is defined as individual scientist's behavior of using other scientists' data for their own research purpose by downloading data from central/local data repositories or requesting the data via personal communication methods. Data reuse does include using the data from the big science research centers for their own research. In this research data reuse was partially considered at the preliminary study; however, the main focus of this research is "data sharing."

1.4. Research Objective and Questions

The main objective of this research is to investigate the factors influencing scientists' data sharing behaviors in diverse scientific communities. This research focuses on scientists' data sharing behaviors, in order to foster data sharing in scientific communities, and eventually help scientists to achieve a core vision of data-intensive scientific research. In order to achieve this goal, this research will have a systematic investigation on the topic area.

This research assumes that scientists' data sharing behaviors are not a matter of individual scientist's arbitrary choice, but rather, decisions on whether to share data with the researchers outside of their research group reflect the choices among communities of colleagues embedded within their disciplines. Therefore, 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.

This research considers the disciplinary differences in scientists' data sharing behaviors as well as individual differences. Since data sharing practices vary depending on scientific disciplines as well as individual scientists (Borgman 2007; Pryor 2009; Tenopir et al. 2011), it is important to understand both disciplinary and individual level factors influencing scientists' data sharing behaviors in diverse scientific communities. There are two primary research questions (RQ) this research aims to address:

> RQ1: What are the institutional and individual factors that influence scientists' data sharing behaviors?
>
> RQ2: To what extent do those factors influence scientists' data sharing behaviors in diverse disciplines?

The first research question aims to identify both institutional and individual factors that influence scientists' data sharing behaviors in general. The preliminary study of this research exactly covers the first research question by exploring the factors motivating and discouraging scientists' current data sharing behaviors. The second research question aims to investigate the extent to which institutional and individual factors identified at the previous stage influence scientists' data sharing behaviors in general. Survey method was used to test the research model with scientists in diverse disciplines. These two research questions are interconnected, and by addressing those two research questions, this research can provide a refined view of scientists' data sharing behaviors across disciplines.

1.5. Theoretical Perspective

Contemporary collaboration in science and engineering fields requires the orchestration of technological infrastructure, institutional support, and interpersonal interactions (Kim et al. 2012). Similarly, scientists' data sharing as the microcosm of contemporary collaboration involves the same three areas of infrastructure, institutions, and people. Individual scientists are nested in institutional contexts, including belonging to universities and academic disciplines, and support from organizational and disciplinary technological infrastructure. In order to understand scientists' data sharing behaviors, this research considers the combination of infrastructure, institution, and people as important components influencing scientists' data sharing.

For example, a scientific discipline may have well-established data sharing practices supported by infrastructure, institutions, and scientists inside the discipline. Scientists' data sharing is facilitated through disciplinary data repositories (as technological infrastructure) where the scientists, in the discipline, can upload their own research data and download others' data. Also, the repositories are made available by organizational support. In addition, the discipline may have strong institutional support, which encourages scientists to share data. Institutional support may include requirements by funding agencies and journals, professional associations' pressures, and the discipline's norms about data sharing. Lastly, individual scientist's perceptions and attitudes toward data sharing may also interact with both infrastructures and diverse institutions and therefore, scientists nested in their institutions are influenced by these technological resources. Scientists actively interacting with their organizations and disciplines will eventually make their own decisions on data sharing behaviors.

In order to fully understand scientists' data sharing, we need to consider how individual scientists make their decisions under institutional influences. In the pursuit of data sharing by an individual scientist, how the institution is set up may influence an individual scientist's decision making. Although some institutions provide a well-designed institutional repository and have some institutional requirements for their scientists to share data, scientists also need to see personal and/or professional value in sharing data in institutional repositories. In other words, scientists make their decisions in the context of belonging to universities, professional associations, academic disciplines, journals, and funding agencies when deciding to share their data with others. At the same time, individual scientists need to have information and technology management skills to prepare and submit the data. Any human and IT support (training) by their affiliated organizations can reduce the barriers involved in data sharing. Therefore, individual scientist's decision making toward data sharing must also be understood within the institutional contexts and technological infrastructure, which are inter-connected. Institutional theory is a perspective from sociology and organizational studies that may help to weave together the intertwined forces of institutions, infrastructure, and people. Institutional theory can provide insight about how social actors are influenced by institutional pressures from the institutional environment. While the traditional focus of institutional theory was at the organizational level of analysis, neo-institutional theory extends its scope to diverse social actors, including individuals as well as organizations under their institutional contexts (Scott 2001). The neo-institutional theory assumes that institutional environments including institutional rules, norms, and culture influence individuals' attitudes and behaviors (George et al. 2006; Tolbert 1985; Tolbert et al. 1983).

Contemporary perspectives on institutional theory consider individual beliefs concerning proper social behavior and, specifically, when those beliefs arise from organizational rules, structures, and practices (Barley et al. 1997; Daniels et al. 2002; Duxbury et al. 1991). This connects nicely with individual-level motivation theories, which describe individual behavior as jointly influenced by beliefs, attitudes, norms, and intentions. This study employs the theory of planned behavior as an individual motivational theory, which can then be connected with institutional theory. The theory of planned behavior provides insights regarding how individuals' attitude, subjective norm, and perceived behavioral control influences individuals' behaviors mediated by intention. The integration of institutional theory and theory of planned behavior can both better explain scientists' motivations and how seeking organizational legitimacy is influenced by institutional pressures. This study can help to validate new theoretical frameworks of the combination of institutional theory and the theory of planned behavior.

1.6. Significance of the Study

This research is significant in terms of theory, method, research (field), and practice. In the theoretical perspective, the integration of institutional theory and individual motivation theory (i.e. theory of planned behavior) can provide a new theoretical lens to understanding scientists' data sharing behaviors. The theoretical framework can offer an insight into how institutional and individual factors influence scientists' data sharing behaviors together. Furthermore, this research can show how individuals' beliefs, attitudes, and behaviors are influenced by and constituted by institutional contexts, and how these institutional influences can be interpreted differently according to individuals' motivations. In terms of theoretical contribution, this research can link the micro level theory that examines individuals' motivations with the macro level theory that examines the role of institutional influences.

Although neo-institutional theory considers individuals' attitudes and behaviors in the context of their institution, not many studies have been conducted which empirically explain the mechanism of how institutions actually influence individuals' behaviors (or intentions). By empirical study, this research can help to validate the main assumptions of neo-institutional theory, or that institutional pressures (logics) affect individuals' attitudes and behaviors. This study can bridge the gap between the neo-institutional theory's perspective and the psychological explanation of attitude and behavior by theory of planned behavior. In addition, by considering the context of institution, this study can make progress in the field of theory of planned behavior, which uses the de-contextualized model of individual level analyses.

In the methodological perspective, this research employs a mixed-method approach with multilevel analysis, and with extensive triangulation can help to understand the phenomena of scientists' data sharing. The mixed-method of combining qualitative and quantitative approaches can provide more fruitful outcomes in studying scientists' data sharing behaviors. Since prior studies have not been conducted in this area and because of the complex nature of data sharing in different scientific communities, the mixed-method approach should be useful. In addition, this research employs a multilevel analysis to investigate the factors influencing scientists' data sharing behaviors at both discipline and individual levels. The employment of multilevel analysis can disentangle the dynamics of institutional and individual effects on scientists' data sharing behaviors.

In the research (field) perspective, this research can provide valuable insights into the domains of scholarly communication and data curation. The advancement of information technologies changed the way scientists communicate and collaborate regarding their scholarly works from traditional publications or article pre-prints to original data. This research can contribute to the area of scholarly communication by examining scientists' emerging scientific communication methods based on their original data. In addition, understanding data sharing is important for library and data curation. Libraries and librarians can provide their expertise and systems for scientists' data curation, and therefore facilitate their data sharing and reuse (Borgman 2010). Delserone (2008) emphasized "data service" as being one of the core services and areas of expertise in library services, and it will potentially support e-Science by building knowledge and capacity within the libraries. By understanding the nature of scientists' data sharing, this research can provide valuable insights for data curation in terms of how to provide any necessary service to help scientists to share and reuse data.

In the practical perspective, this research can help scientific communities by possibly accelerating scientists' data sharing behaviors as a part of their scientific collaborations, and eventually enable the vision of data-intensive scientific research. By understanding scientists' data sharing in the institutional and individual perspectives, this research can provide useful guidelines and recommendations in designing metadata standards and repositories. Also, this research can help to develop relevant policies for data sharing which best facilitate individual scientists' data sharing in different scientific communities.

First, the effective development of data repositories requires the careful understanding of scientists' data sharing practices. Borgman and colleagues (2007a) also argued that the design and development of data repositories and information services need to consider data practices in their user communities. The final outcomes of this research can provide valuable insights to better guide the development of central or local data repositories in different disciplines. This research can examine the roles of metadata and data repository in regards to scientists' data sharing, and it can help scientific communities to manage their existing or future metadata and repositories to best facilitate scientists' data sharing.

Second, this research can also provide valuable insights for designing relevant policies for data sharing in the perspectives of funding agencies and journal publishers. Many journals in science and engineering research now require that their authors submit the experiment's data to relevant data repositories and/or provide their data to other scientists upon request. Recently, national and public funding agencies have required their grant awardees to share the primary data with other scientists as a part of their data management requirements. However, the effectiveness of these policies toward scientists' data sharing is still in question. This research can show how institutional policies, such as those of funding agencies and journals, are influencing scientists' data sharing.

1.7. Summary

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 and reusing others' data through technological development and collaborative effort (Hey et al. 2008).

In the perspective of scholarly communication, primary data collected by individual scientists becomes an important “information currency” along with research analyses and finding in the traditional publications (Davis et al. 2007). As the primary data becomes important in terms of data-intensive scientific research and scholarly communication, data sharing practices are now essential in most modern research activities.

The objective of this research is to investigate the factors influencing scientists’ data sharing behaviors in different scientific communities by examining both discipline and individual level predictors together. Since data sharing varies depending on scientific disciplines (Borgman 2007; Pryor 2009; Tenopir et al. 2011), it is important to explore the institutional factors as well as individual factors influencing scientists’ data sharing behaviors across various scientific communities. In summary, both institutional and individual factors influencing scientists’ data sharing behaviors need to be examined carefully. This investigation can provide a holistic view of institutional and individual factors influencing scientists’ data sharing across diverse scientific disciplines.

This research is significant in terms of theory, method, research, and practice. In the theoretical perspective, the integration of institutional theory and individual motivation theory (i.e. theory of planned behavior) can provide a new theoretical lens to understanding scientists’ data sharing behaviors. In the methodological perspective, this research employs a mixed-method approach with multilevel analysis and with extensive triangulation. In the research perspective, this research can provide valuable insights to the domains of scholarly communication and data curation. In the practical perspective, this research can help scientific communities by possibly accelerating scientists’ data sharing as a part of their scientific collaborations, and eventually enable the vision of data-intensive scientific research.

2. Literature Review

This chapter reviews scientists' norms and values, scholarly communication, and the literature of scientists' data sharing and reuse. In order to understand scientists' data sharing behaviors, this research considers scientists' norms and values as the structure of science. Also, this research provides the overview of scholarly communication in regards to data sharing. Then, the synthesis of the literature on scientists' data sharing/withholding and reuse is provided. Prior studies in data sharing/withholding have focused on prevalence of data sharing/withholding, factors influencing data sharing/withholding, and the consequences of data sharing/withholding. Lastly, this chapter provides the limitations of previous studies.

2.1. Scientists' Norms and Values

In order to understand scientists' data sharing behaviors, this research considers scientists' norms and values as the structure of science. Scientific norms and values are embedded in scientists' data sharing behaviors as seen in scholarly communications. According to Robert K. Merton's (1973) research, science's norms and value system make science different from other social institutions. This section reviews the nature of science, scientists' norms and values, and scholarly communication as the basis of data sharing.

2.1.1. Nature of Science

In order to study scientists' data sharing, it is important to understand the nature of science. Scientists conduct research by stating research problems, acknowledging previous literature, conducting research, interpreting research findings, and using publication channels (Pierce 1990). Popper (1968) and Gauch (2003) identified several additional steps of scientific research, including (1) observing and experiencing natural/social phenomena, (2) developing hypotheses and predictions, (3) testing those hypotheses and predictions, and (4) presenting findings and deriving conclusions which may generate new hypotheses or refute the old ones.

Because the scientific research process relies entirely on evidence and logic, science is generally assumed to produce superior knowledge (Merton 1973). Scientific research uses science's own methods, standards, norms, and mechanisms to generate and evaluate knowledge. In particular, scientific research has developed through the diverse scholarly communication mechanisms of peer-review, publication, citation, and criticism for validity and further research. All of these norms and mechanisms to facilitate the production of scientific knowledge enhance scientific superiority and make science an institution able to exist by itself with a self-controlling system. Merton (1973) also argued that scientific superiority has been enhanced by following its scientific methods and norms which facilitate the production of scientific knowledge.

Science is considered as both an autonomous and a social institution, which is to say, both independent from and dependent upon other institutions. Science as an institution is free from external controls and judgments, which means that the scientific community has the right to self-control its own research activities by leading its own research agendas and evaluating research findings in its knowledge production (Barber 1952; Goldsmith 1967; Merton 1970; Polanyi 1945; Richter 1980). At the same time, science is also a social institution. Scientists and other institutions interact and have a close relationship with society (Ziman 2000). Merton (1973) argued that it is important to avoid the simplistic view that science is autonomous and independent from external controls and judgments. Science is embedded in its social and cultural contexts, and is usually influenced by the economy, culture, and other external forces (Bloor 1976; Pinch et al. 1984). Whitley (2000) indicated that science as a social institution has changed its structure and operation through industrialization by depending on other social institutions (p. 266). Additionally, the scientific community relies on the government and other organizations for its funding (McGrath 2002). Since the relationship between scientists and funders is hierarchical, their scientific research may be influenced by funders.

2.1.2. Norms of Science

Science as a social activity relies on interaction between individual scientists (Kuhn 1996). Social interaction within the scientific communities follows the norms that regulate scientific research, practice, publication, and scientists' data sharing practices. Understanding scientific norms is important because the norms would influence scientists' data sharing practices. Scientists conform to these community standards because they make scientific research more valid and reliable. There are no absolute norms that affect scientists across time and scientific disciplines; however, there are both traditional norms and counter-norms that affect social practices in different scientific communities. Merton (1973) defined the four traditional norms of science as communalism, universalism, disinterestedness, and organized skepticism. These scientific norms can explain how a scientific community works.

Communalism

Communalism in this context means that scientific findings must be made available to the general public and shared with all members of the scientific community (Braxton 1986). Merton (1973) argued that scientific findings should be owned by the community that produces them, because most scientific findings are based on collaboration among scholars and on the work of previous scholars. If scientists provide the scientific community with insufficient information about their findings, other scientists will be unable to replicate or disprove the original findings. Communalism enables open and free sharing of scientific knowledge, and it also encourages the sharing of supporting data along with the final analysis and results.

Universalism

Universalism in the scientific community means that scientific research must be judged by scientific criteria rather than by identities of the scientists (Merton 1973). This norm tells that research needs to be judged by the standardized criteria of research rather than scientists' diverse social characteristics, including scientists' race, gender, class, religion, and other personal characteristics. Universalism employs universal criteria to generate, manage, and evaluate knowledge (Merton 1973). The blind review process in peer-reviewed journals is a good example of the idea of universalism in practice. Gaston (1973) found that social class origins and educational backgrounds do not significantly influence scientists' research productivity in England's high-energy physics community.

Disinterestedness

Disinterestedness is defined as “the preference for the advancement of knowledge as opposed to the individual motives of the scientist” (Braxton 1986). Disinterestedness means that scientists must be detached from their personal economic rewards toward their research. According to the norm of disinterestedness, scientists are supposed to be less interested in any personal reward (e.g. financial benefits or personal reputation) for their research than in the development of scientific knowledge in their research community. Disinterestedness also prohibits scientists from aligning their research with funding opportunities (MacFarlane et al. 2008).

Organized Skepticism

Organized skepticism as a scientific community norm means that scientific findings should be examined for empirical evidence of scientific merit before being accepted as new scientific knowledge (Merton 1973). According to this norm, scientists must review all scientific findings with a degree of skepticism, even their own research findings (Merton 1973). Published scientific work must be possible to replicate; if it is not, it must be denied. All these conditions must hold before findings can be accepted; scientists can replicate or deny any scientific work which was published for the public based on the norm of organized skepticism. Organized skepticism requires scientists to examine other scientists’ works in terms of empirical evidence and logics before they accept the findings as true scientific knowledge.

2.1.3. Counter Norms

Other scholars reconsidered the scientists' conformity to Merton's four scientific norms arguing that scientists do not behave entirely according to Merton's four scientific norms, but rather also seek their own interests through scientific research (Mitroff 1974; Mulkay 1976). Mulkay (1976) found that scientists use scientific norms to negotiate and justify self-interested behaviors in relation to scientific norms and their interests. Mitroff (1974) provided counter-norms to Merton's four norms of science, including solitariness, particularism, interestedness, and organized dogmatism. Mitroff (1974) argued that Merton's original norms of science and his alternative norms are mixed in an actual science institution.

Solitariness

Solitariness as the counter-norm to communalism means that scientists consider their research findings as protected property and feel secrecy is needed to protect their rights over their research findings (Mitroff 1974). Scientists are also more interested in their intellectual property than in project completion and publication (Brown 2003; Marshall 1990; McCain 1991). Under solitariness, scientific findings belong to the scientists who identify those findings, not to the whole scientific community (Mitroff 1974). Those scientists will protect their research findings with patents or property rights. Scientists' funding sources also encourage solitariness. Research studies funded by private companies or organizations become secret to other members of the scientific community (Mowery 2005).

Particularism

Particularism, as the counter-norm to universalism, is judging scientific findings according to scientists' social backgrounds (Mitroff 1974). Mulkay (1976) argued that the distribution of recognition is biased toward researchers of prestigious universities. Merton and Sztompka (1996) identified another example of particularism, known as the Matthew Effect. That example holds that the scientist who made a valuable scientific finding can be considered as having more merit because of his or her reputation. According to the Matthew Effect, scientists who make a significant scientific improvement in their disciplines tend to have more unquestioned credibility than they should have (Merton et al. 1996). The blanket acceptance of scientific findings by well-known scholars is another example of particularism (Andersen 2001).

Interestedness

Interestedness is the counter-norm of disinterestedness, and it means that scientists care more about personal financial benefits from research than about personal satisfaction and reputation from scientific findings (Mitroff 1974). According to interestedness, scientists seek personal financial rewards through their research performance. Another form of interestedness is developing a research agenda based on funding opportunities, rather than on a desire to seek scientific findings in the scientist's area of research interest.

Organized Dogmatism

Organized dogmatism as the counter-norm to organized skepticism means that scientists accept certain scientific findings without examining them carefully (Mitroff 1974). Scientists need to be skeptical of previous findings before they accept them as new scientific knowledge. Scientists need to indicate the shortcomings of previous research when their research findings invalidate the earlier studies (Mitroff 1974). Another form of organized dogmatism occurs when a scientist is skeptical of other scientists' findings (but not his or her own), although the scientist needs to be skeptical of his or her findings as well as of others' findings (Mitroff 1974). Table 2.1 below shows a summary of Merton' (1973) norms of science and Mitroff's (1974) counter norms:

Definitions	Norms of Science	Counter Norms	Definitions
Scientific findings must be shared with all members of the scientific community	Communalism	Solitariness	Scientists consider research findings as protected property and secrecy is needed to protect them
Scientific research must be judged by scientific criteria rather than by scientists	Universalism	Particularism	Judging scientific findings according to scientists' social backgrounds
The preference for the advancement of knowledge as opposed to the individual motives of the scientist	Disinterestedness	Interestedness	Scientists care more about financial benefits from research than personal satisfaction and reputation
Scientific findings should be examined for empirical evidence of scientific merit before being accepted	Organized Skepticism	Organized Dogmatism	Scientists accept certain scientific findings without examining them carefully

Table 2.1 Summary of Merton' (1973) Norms of Science and Mitroff's (1974) Counter Norms

2.1.4. Values of Science

Merton (1957) described the race for priority, which showed that scientists place high value on being recognized as the first discoverer of scientific findings. Academic reputation based on research production is an important value in many scientific communities (Merton 1957). The motivation for scientists to achieve reputation is an important value in many scientific communities. Merton (1973) argued that scientific institutions work based on a reward system in which recognition and credit go to those who make original contributions to scientific knowledge. Many previous studies have found that scientists work based on the reward of a favorable reputation (Dundar et al. 1998).

Scientists internalize the four scientific norms as institutionalized values (Merton 1973). The value of credit and priority in scientific communities supports Merton's four norms of communalism, universalism, disinterestedness, and organized skepticism (Merton 1957). Scientists can gain rewards in the form of reputation and credit by sharing their research findings with other scientists without any limitation (communalism). Their reputations and rewards are not supposed to be based on their social or educational backgrounds, but rather on the quality of their research (universalism) (Cole et al. 1973). Scientists want to be recognized for advancing science knowledge rather than being satisfied with monetary benefits (disinterestedness). Lastly, the scientific community is supposed to provide appropriate credit to scientists who contributed to knowledge of science only after members of the community examine previous studies in terms of their empirical evidences and logic (organized skepticism).

The reward system in science is associated with the publication of research as a scholarly communication practice (Borgman 2007). Scientists achieve science's core values by publishing and being cited by other scientists. As Latour (1987) indicated, citations can provide justifications and appropriate rewards for scientists' research findings. Through citations, scientists acknowledge previous research and provide appropriate credit.

Scientists would conform to the norms of science and to science's institutionalized reward system in order to achieve their values in science. Publication as scholarly communication practice supports the scientific community's reward systems. Scientists' data sharing is, like publication, an extension of scientists' scholarly communication. As such, the norms and values of science can be applied to scientists' data sharing practices. Previous studies about scientists' data sharing show the coexistence of Merton's traditional norms of science and Mitroff's counter norms of science (Louis et al. 2002). For example, McCain's (1991) study found that geneticists behave based on communalism and disinterestedness; however, Ceci's (1988) study showed that geneticists follow solitariness and interestedness as norms.

2.2. Scholarly Communication

Conducting scientific research requires salient communication features for sharing scientific findings and knowledge (Garvey 1979). In the field of information science, these communication features are called scholarly communication, which is "the study of how scholars in any field use and disseminate information through formal and informal channels" (Borgman 1990). Scientists generate scientific findings and knowledge, and they disseminate and discuss scientific findings and knowledge through diverse formal communication channels such as journals and conferences (Pierce 1990). Additionally, they use interpersonal networks to discuss and disseminate research findings and scientific knowledge through information communication channels such as personal electronic communications.

Formal scientific communication channels establish the priority of scientific research findings and support reward systems in scientific communities (Zimmerman 2003). As a major formal scientific communication channel, peer-reviewed journals work as a window for disseminating and evaluating scientific knowledge (Schickore 2008). Journals also facilitate scientific communication by helping scientists share and discuss their research findings. More importantly, the system of journal publication supports the values of science by providing scientists with appropriate rewards (i.e. priority and credit) through the mechanism of publication and citation.

Traditionally, formal scholarly communication is based on journal articles and conference proceedings; more recently, it has also been based on article preprints in some disciplines. Scientists share their knowledge through these formal communication channels by locating relevant information from articles. However, modern scientific research requires original data sets for diverse purposes such as large-scale computation, comparative research, or replication of previous works for further research. Borgman (2007) also argued that sharing data as well as publishing improves scientific communication by increasing research transparency and reproducibility. For example, many research works in the field of biology require data collected by other scientists in order to validate previous research, and future research is often designed to duplicate previous works. Therefore, in the perspective of scholarly communication, data sharing becomes important in modern scientific research activities (Cragin et al. 2006).

Emerging information and communication technologies have enabled new scholarly communication methods of data sharing. Scientists share data through informal communication channels such as email, Web file sharing, and FTP services, and share data through formal communication channels such as local or central data repositories. Khatibi and Montazer (2009) argued that electronic scientific databases (i.e. data repositories) enhance scholars' research processes by facilitating scientific communications and collaborations based on the original data sets.

Scholars treat both research publications and data as important sources of scholarly communication. Raw data sets have become important "information currency" for scholarly communication, as they supplement traditional research analysis and findings in journal publications (Davis et al. 2007). Although data sets have become an important form of scientific communication, there have been few studies on how scientists' data collection, management, analysis, and archiving support scholarly communications (Heidorn 2008). Understanding scientists' data sharing behaviors can help scientific communication scholars to better support scientists in their data management and in their scientific research.

Sharing data among scientists means that more scientists can benefit from the data; however, data sharing has not yet been established as major scholarly communication methods throughout different scientific communities (Borgman 2007). Rather, each discipline has developed its own informal or formal data sharing practices associated with its scholarly communication practices. The new system of scientific communication takes a long time to fulfill the emerging need for reliable transfer of scientific knowledge (Zimmerman 2003). Data sharing would be desirable scientific behavior under the norms of communalism and disinterestedness. However, 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). A standardized citation system would help scientists to achieve their values. This research involves diverse issues of data sharing as a new method of scientific communication. The next section will discuss in detail issues of data sharing and reuse in previous literature.

2.3. Data Sharing/Withholding

This literature review covers prior studies in not only scientists' data sharing behaviors, but also their data withholding behaviors. This research focuses on data sharing behavior, which means providing raw data to other scientists by making it accessible through data repositories, or by sending data via personal communication methods upon request. In the literature review, data withholding behavior as an opposite form of data sharing behavior was considered. Data withholding behavior can be defined as refusing to provide raw data to other scientists when scientists are expected to provide their data by depositing it into data repositories, or by sending it via personal communication methods upon request.

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 (other) consequences of data sharing and withholding (Campbell et al. 2002; Campbell et al. 1998; Campbell et al. 2000; Louis et al. 2002). Although data sharing is desirable according to scientific communities' norms of communalism and disinterestedness and can contribute to the advancement of scientific research, there is ample evidence that scientists nonetheless withhold their data rather than sharing it in popular science journals (Campbell et al. 2003; Cohen 1995; Piwowar 2011). A good amount of previous data sharing research has focused on whether scientists allow or deny other researchers access to their data (e.g. Campbell et al. 2002; McCain 1991).

2.3.1. Prevalence of Data Sharing/Withholding

Most previous research on data sharing and withholding has studied the prevalence of data withholding rather than data sharing. Many such studies have focused on one specific form of data withholding: scientists' denial of others' requests for the raw data used in their published research (Campbell et al. 2002). Blumenthal and colleagues (1997) surveyed life scientists across the nation and discovered that 8.9% of those life science researchers had denied a request for the data used in their publications. A later study by Campbell and colleagues (2002) found that during the previous three years, 12% of geneticists at U.S. major research universities had denied other researchers access to their publication related information. Vogeli and colleagues (2006) reported that 7.9% of science trainees had denied other researchers' requests to access the data for their own published research. Another study, with faculty members at U.S. medical schools, found a slightly higher 12.5% denied request rate between 1996 and 1997 (within last three years) (Campbell et al. 2000).

Data withholding rates vary across different disciplines and through different publication stages (Borgman 2007; Pryor 2009; Tenopir et al. 2011). For example, Blumenthal and colleagues (1997) found that geneticists in the field of life science were more likely to deny others' requests than were non-geneticists in that field. Blumenthal and colleagues (2006) confirmed this finding in 2000 by surveying U.S. geneticists and other life scientists, where 44% of geneticists and 32% of other life scientists participated in various forms of data withholding during the three years prior to the study.

Data withholding rates also depend on the publication status of research. In another study of geneticists, Louis and colleagues (2002) found that 30% of genetic researchers reported that they withheld data at least once, pre-publication, within the past three years. Vogeli and colleagues (2006) surveyed science trainees regarding data withholding and found that 23.0% of trainees were denied access to publication related materials and 20.6% were denied access to unpublished research. Similarly, Blumenthal and colleagues (2006) found that data about published articles was more often withheld (geneticists 35%, other life scientists 25%) than was the data about pre-published works (geneticists 23%, other life scientists 12%).

Data withholding behaviors also vary from discipline to discipline. Reidpath and Allotey (2001) requested publication-related data from the authors of 29 articles published in the British Medical Journal. Only one author released the data requested. In another behavioral research study, Savage and Vickers (2009) requested data sets from the authors of 10 articles published in the PLoS (Public Library of Science) journals, which represent the new trend of “open access”, and received only one response.

Studies related to data sharing often use bibliometrics analysis to explain data sharing’s prevalence. One such study by Piwowar and Chapman (2008b) investigated the prevalence of data sharing regarding gene expression microarray data by counting the papers that linked to NCBI (National Center for Biotechnology Information)’s Gene Expression Omnibus (GEO) database. More recently, Piwowar (2011) conducted another study that used bibliometric analysis to identify how frequently raw gene expression microarray datasets were shared after publication. She found that 25% of the 11,603 articles about gene expression microarray published between 2000 and 2009 provided their raw datasets in major data repositories. This shows that the actual rate of data sharing within the scientific community is relatively low (Blumenthal et al. 2006), and it varies by discipline (Borgman 2007; Pryor 2009; Tenopir et al. 2011).

In the field of psychology, Wicherts and colleagues (2006) requested research-related information from 141 authors of articles published in American Psychological Association (APA) journals. They found that only 38, or 27.0%, of those authors released research-related data upon request. This response rate is similar to the response rate of 24.3% (9 out of 37 requests) which Wolins (1962) reported when they requested data from 37 authors who published articles in APA journals. Similarly, Craig and Reese (1973) reported that 37.7% of authors (20 out of 53) provided either original data or a summary of data analysis in major APA journals. These studies show that data sharing varies by discipline, and that in the field of psychology, the data-sharing rate has decreased over the past several decades, despite advances in technological communication tools and the widespread availability of the Internet. Table 2.2 below shows the summary of prior research findings about prevalence of data withholding:

Withholding Types	Sources	Subject/Discipline	Withholding Rate
Denying a request for the data of published articles	(Blumenthal et al. 1997)	Life Scientists	8.9%
	(Campbell et al. 2000)	Medical Scientists	12.5%
	(Campbell et al. 2002)	Geneticists	12.0%
	(Vogeli et al. 2006)	Science Trainees	7.9%

Denying a request for the data of published and unpublished works	(Louis et al. 2002)	Geneticists	30%
	(Vogeli et al. 2006)	Science Trainees	20.6%
	(Blumenthal et al. 2006)	Geneticists	23%
	(Blumenthal et al. 2006)	Other Life Scientists	12%
Withholding data in various forms	(Blumenthal et al. 2006)	Geneticists	44%
	(Blumenthal et al. 2006)	Other Life Scientists	32%
Experiment study by requesting the data of published articles	(Wolins 1962)	American Psychological Association Journals	24.3% (9/37)
	(Craig et al. 1973)		37.7% (20/53)
	(Wicherts et al. 2006)		27.0% (38/141)
	(Reidpath et al. 2001)	British Medical Journal	3.4% (1/29)
	(Savage et al. 2009)	PLoS Journals	10% (1/10)
Depositing the gene expression microarray	(Piwowar 2011)	Geneticists (Microarray)	25% (Sharing Rate)

Table 2.2 Summary of Prior Research Findings about Prevalence of Data Withholding

2.3.2. Factors Influencing Data Sharing/Withholding

Prior studies provide research on diverse factors influencing scientists' data sharing and withholding. According to the theoretical perspective considering the combination of institution, infrastructure, and people as important components influencing scientists' data sharing behaviors, I categorized those factors into three groups. These include: institutional factors (i.e. funding agency's policy, journal requirements, and contract with industry sponsors); resource factors (i.e. metadata and data repositories); and individual factors (i.e. personal characteristics, perceived benefit, perceived effort, perceived risk).

In addition, other organizational and environmental factors have been studied as important factors influencing scientists' data sharing and withholding.

Institutional Factors

Funding Agency's Policy

Stanley and Stanley (1988) argued that contemporary scientists consider data sharing among researchers to be an obligation rather than a voluntary activity. Funding agencies' policies help to cause this sense of obligation. Scientific funding agencies such as National Institute of Health (NIH) and National Science Foundation (NSF) require their grant awardees to allow shared access to the data collected (National Institutes of Health 2003; National Science Foundation 2010). Scientific organizations across a variety of disciplines have implemented similar policies mandating data sharing (Faniel et al. 2011). Scientific communities are gradually agreeing that research data generated using public funding needs to be freely and openly available to all interested parties (Arzberger et al. 2004).

Since 2003, the NIH in the U.S. has required any project that receives more than $500,000 of funding per year follow the NIH's data sharing policies (National Institutes of Health 2003), and the NSF recently mandated that grant awardees make a data management plan as a condition of their funding (National Science Foundation 2010). The National Cancer Institute (NCI) also requires grant applicants to create a data sharing plan (Colditz 2009).

Researchers studied the correlation between the data sharing policy and the scientists' data sharing; scholars found that these data sharing policies caused community pressure to share scientific data (McCullough et al. 2008; Piwowar et al. 2008a). Similarly, based on bibliometric analysis, Piwowar and Chapman (2008b) found a significant correlation between funding agencies' data sharing requirements and scientists' actual data sharing. Still another study found that scientists who received a large number of NIH grants were more likely to share their data with others (Piwowar 2011). However, Piwowar and Chapman (2010) found that there was no significant correlation between the NIH data sharing requirement and scientists' actual data sharing behavior. According to their findings, data sharing had not significantly increased over the last 10 years. Studies on funding agencies' data sharing policies and their influence on scientists' data sharing often draw mixed or contradictory conclusions, and have focused on specific subgroups of scientists rather than scientists as a whole.

Journal Requirements

Just as funding agencies created their own data sharing policies, journals have implemented their own data sharing policies affecting the scientists whose articles they publish (McCain 1995; Piwowar et al. 2008a; Piwowar et al. 2008b). McCain (1995) found that only 132 out of 850 natural science, medical, and engineering journals had at least one journal policy statement mandating (1) depositing data in publicly available data repositories, (2) sharing research related materials upon request, and (3) providing supplementary publication-related services. However, now many biomedicine and molecular biology journals require scientists to submit original datasets to databases once their articles are accepted (Brown 2003; McCain 1995; Piwowar et al. 2008a; Piwowar et al. 2008b). Bebeau and Monson (2011) found that social science fields such as psychology, sociology, and education also have data sharing agreements in the form of ethics. A recent study by Piwowar and Chapman (2008b) reviewed 70 journal policies in the research area of microarrays and found that 52 out of the 70 journals, or 74.3%, explicitly mentioned data sharing requirements. Many journals now require authors to share information with other researchers either by depositing their data in publically available data repositories or by providing the data freely upon request (Savage et al. 2009).

Several studies have tested the relationship between journals' data sharing policies and actual data sharing behavior. Piwowar and Chapman (2010) reviewed the database submission information in the articles published in the journals that required authors to deposit their original data, and observed that studies published in these journals tended to share their data through data repositories. Piwowar and Chapman (2008b) found that there is a positive correlation between the strength of journals' data sharing policies and the rate at which scientists deposit data in a public database. Scholars who published articles in prestigious journals were also more likely to share their data in data repositories (Piwowar et al. 2010), as were the authors of articles published in open access journals (Piwowar 2011).

However, several studies pointed out that, in the actual practice of data sharing in different scientific fields, the publication-related data and materials are not always available for other researchers (Cech et al. 2003). Noor and colleagues (2006) found that for 3% to 20% of articles published in genetics journals with clear data sharing policies, authors did not deposit their data in any relevant data repositories. Another study by Savage and Vickers (2009) investigated whether the authors whose articles were published in journals with strong data sharing policies provided raw datasets when requested; they found that only one author sent data out of the 10 requests made. These studies show that journals' data sharing policies positively influence the prevalence of sharing data through depositing it in central data repositories; however those policies still do not consistently motivate scientists to share their data, through either data repositories or personal communications (e.g. email).

Contract with Industry Sponsors

Industry sponsors are common in many science and engineering fields, and they support a great deal of research. However, previous studies have found that contracts with industry sponsors make scientists less likely to share their data with others (Louis et al. 2002). Campbell and colleagues (1998) found that industry sponsors often place restrictions on the research outcomes supported by their funding, which prevented scientists from sharing data with others. Louis and colleagues (2002) reported that 21% of geneticists withheld their data in order to keep agreements with industry sponsors. Campbell and Bendavid (2003) found that government agencies sometimes provide scientists with funding under strict policies about data sharing, even though these government projects are publicly-funded research. In a recent study, Blumenthal and colleagues (2006) found that geneticists or other life scientists participating in close relationships with industry were more likely to withhold data both verbally and in published form. Additional studies found that faculty members were reluctant to submit data to data repositories for fear of copyright or contract infringement (Foster et al. 2005).

Resource Factors

Metadata Standards

Metadata standard is an important factor in scientists' data sharing. Metadata is defined as data about data that formalizes and standardizes unorganized data (Zimmerman 2007). Standardized data vocabularies help scientists to avoid generating heterogeneous representations of similar datasets (Saltz et al. 2006). The limitations of metadata standards and descriptions makes data sharing more difficult for scientists to discover and use data from more than one research center (Horsburgh et al. 2011). Scholars argued that in order to stabilize and maintain scientific data, scientific researchers must develop consistent metadata standards (Bowker et al. 2000). Recently, many research groups have introduced and encouraged the adoption of metadata standards to enable data discovery and reuse (Bietz et al. 2010; Field et al. 2008; Hey et al. 2004; Karasti et al. 2010).

Previous studies have largely focused on the development of metadata standards within specific scientific fields (Diaz et al. 2011; Karasti et al. 2008; Millerand et al. 2010; Ribes et al. 2010). For example, the field of ecology developed the Ecological Metadata Language (EML) to organize and manage ecological data (Karasti et al. 2008), and the field of life science developed its own metadata standard for experimental research to encourage data sharing and archiving (Paton 2008). Standardized data and metadata allow for a more collective research practice (Ribes et al. 2010) and for data integration in a distributed environment (Diaz et al. 2011). However, most previous studies on metadata standards have focused on data sharing and reuse in research collaboration projects rather than on allowing access to publication data. Therefore, it is necessary to study whether metadata standards can facilitate scientists' data sharing by reducing the time and effort it takes for them to share their data.

Data Repositories

The availability of data repositories can be another important factor affecting scientists' data sharing. Data repositories were designed to allow research communities to store, share, query, and download data (Fennema-Notestine 2009; Horsburgh et al. 2011). They help scientists to validate results, facilitate reuse and reanalysis, and eventually advance scientific findings through large sets of data (Schwartz et al. 2010). There are web-based data repositories available across many different scientific disciplines including biology, genetics, medicine, geosciences, and astronomy (Eschenfelder et al. 2011). Institutional repositories at universities provide additional data management support such as electronic documents, digital archival collections, and data curation (Choudhury 2008; Witt 2008). A well-known example of an institutional data repository is the DataStaR (Data Staging Repository) hosted by Cornell University. The DataStaR is a temporary local data repository designed to support data sharing among research collaborators during the research process and to help scientists publish quality data and metadata in an external repository supported by librarians (Steinhart 2007).

Previous studies have found that both disciplinary and organizational data repositories facilitate and promote scientists' data sharing (Marcial et al. 2010). Brown (2003) argued that in the field of molecular biology, the acceptance and usage of disciplinary data repositories have improved research dramatically, by providing a storage and retrieval mechanism for the research data in the field's publications. Cragin and colleagues (2010) also investigated how institutional data sharing repositories influence scientists' data sharing, and concluded that institutional repositories can facilitate data sharing among scientists by providing data stewardship. They also argued that scientists may have difficulties sharing data in part because data repositories are not readily available or suitable (Cragin et al. 2010). Fennema-Notestine (2009) argued that the Biomedical Data Repository (BDR) in clinical communities has increased data accessibility and supported existing research and education related data sharing structures. However, scholars also argued that scientists do not fully utilize existing data repositories to reuse others' research data (Glover et al. 2006; Karasti et al. 2006).

Individual Factors

Characteristics

Several studies exist on the characteristics of scientists who readily share their data and on the characteristics of scientists who refuse requests for their data (Cragin et al. 2010; Piwowar 2011; Piwowar et al. 2010). Scholars used bibliometric analysis to identify the characteristics of biologists who share their data with others (Piwowar 2011; Piwowar et al. 2010). They found that researchers with high levels of career experience and impact were more likely to share their data (Piwowar et al. 2010), and that the more prior experience authors had with sharing or reusing data, the more likely they were to share their data (Piwowar 2011).

Prior studies found that scientists who deny others' requests for their data have several similar characteristics. Male researchers in particular are more involved in data withholding among geneticists and other life scientists (Blumenthal et al. 2006), and researchers who want to file a patent or commercialize their research results are more likely to withhold and refuse requests for their data (Campbell et al. 2003; Campbell et al. 2002). Campbell and Bendavid (2003) found that 80% of life scientists indicated they needed to keep research results secret for patent filing purposes. In addition, researchers supported by industries are more likely to withhold their data through tactics such as delaying publication by more than six months (Campbell et al. 2003).

In addition, those scientists who “were denied” access to other researchers’ data also have several characteristics in common. Campbell and colleagues (2000) found that scientists who withheld research data, published many articles, and applied for patents were more likely to be refused access to others’ data. Vogeli and colleagues (2006) also argued that scientists are less willing to share their data with those who have industry relationships because of fears that shared data might be used for commercial purposes. However, existing studies on scientists’ characteristics as they relate to data sharing are limited to certain characteristics and specific disciplines; therefore, further research must study a wider range of characteristics within a variety of scientific disciplines.

Perceived Benefits (Reward and Reputation)

Previous studies considered perceived benefit as an important factor influencing scientists’ data sharing. Perceived benefit was studied as a form of reward and reputation in scientists’ data sharing. The reward and reputation of scientific work can be measured by citation counts because the citations are used for research funding, promotion decisions, and salaries, so are a reasonable metric for the perceived benefits of scientific work (Diamond 1986). Previous studies have found that professional recognition (Kim 2007), institutional recognition (Kankanhalli et al. 2005), and academic reward (Kling et al. 2003) influence scientists' data sharing behaviors. Stanley and Stanley (1988) found that when scientists perceive a lack of reputation and recognition incentives in data sharing, they are less likely to share their data. Similarly, Sterling and Weinkam (1990) indicated that the potential loss of monetary, political or psychological reward is one reason scientists do not share their data. However, Piwowar, Day & Fridsma (2007) found that, counter to these scientists' expectations, the number of times a work is cited is positively associated with the public availability of that work's original data. Works containing data available through public data repositories were 69% more likely to be cited (Piwowar et al. 2007).

Reciprocal benefit as a part of perceived benefit was studied as an important factor for internal data sharing (personal data sharing). Social exchange between data producers and reusers, especially as it pertains to perceived reciprocity, influences both scientists' data sharing (Collins 1992) and their knowledge sharing (Nahapiet et al. 1998). One study indicated that scientists share their data among close associates or their own social acquaintances (Zimmerman 2007) because these associates are then more likely to share their own data. Louis and colleagues (2002) found that scientists (28% of geneticists) are reluctant to share their data because others may not reciprocate. However, in the context of modern data sharing, the concept of social exchange may not apply, since scientists provide their data through data repositories or to strangers upon request.

Perceived Efforts (Arrangement and Interoperability)

The time and effort which researchers need to spend are an important factor preventing data sharing. Previous studies on scientists' data sharing have reported that the effort of data sharing, such as organizing and preparing data, prevents scientists from sharing their data with others. Stanley and Stanley (1988) noted that the time and effort which takes to organize or prepare data are critical factors preventing data sharing. According to Campbell and colleagues' (2002) study, 80% of geneticists who denied others' requests reported that they withhold their data simply because producing the publication-related information and data takes too much effort. Louis and colleagues (2002) also noted that more than two-thirds of geneticists were less likely to share prepublication results because of the extra effort involved in sharing data. Foster and Gibbons (2005) and Kim (2007) found that faculty members were reluctant to submit content to institutional repositories because it requires additional work, such as creating metadata. In a recent study, Tenopir and colleagues (2011) found that scientists do not make their data available online because they lack the time and funding to organize their data.

Data sharing requires considerable administrative work, but many scientists do not have enough time and support from their organizations to manage their data (Tenopir et al. 2011). For this reason, scientists may fear information requests because scientists must then spend a significant amount of time addressing those requests (Piwowar 2010). Brandt (2007) argued that scientists do not have time to organize data, so they need institutional support to describe and organize their data for future reuse. Similarly, Giffels (2010) argued that scientists need information experts' support to participate in data sharing because external support is very limited.

Technical issues regarding compatibility and interoperability contribute to the perceived effort involved in data sharing. As modern science becomes more data-driven, collaborative, and interdisciplinary, the interoperability of data and tools becomes increasingly important (Edwards et al. 2011). In particular, the interoperability of technologies is crucial in allowing scientists to collaborate with others in different disciplines (Stein 2008). Previous literature has paid considerable attention to technical aspects of data sharing (Akbulut-Bailey 2011; Arzberger et al. 2004). Several studies have concluded that scientists find data sharing or reuse more difficult and time-consuming if data types and relevant technologies are incompatible or not interoperable (Reitsma et al. 2009).

Perceived Risks (Control, Misuse, Criticism, and Data Sensitivity)

Scientists may view data sharing as risk, which includes losing publication and commercialization opportunities and worrying about misuse and criticism by other scientists. First, one of the main reasons scientists do not want to share their data is that they view data sharing as losing publication opportunities. Scholars found that scientists are reluctant to share their data because of concerns about losing publication opportunities and the exclusive rights to their data (Reidpath et al. 2001; Savage et al. 2009; Stanley et al. 1988). Stanley and Stanley (1988) found that scientists are also concerned about reusers' qualifications and about publicly available data being misused. Louis and colleagues (2002) found that scientists avoid sharing their data in order to protect their own or their students' abilities to publish. Similarly, Campbell and colleagues (2002) reported that geneticists deliberately withhold publication-related data because they want to keep further publication opportunities open for themselves, their graduate students, and postdoctoral fellows. Scientists worried that if they share their data openly, other scientists would be able to publish before they could (Sedberry et al. 2011), so they viewed data sharing as losing future opportunities to improve their reputations and receive other benefits of publication (Walsh et al. 2003). Weil and Hollander (1991) described this pattern as the desire to protect scientists' scientific priority.

Additionally, the trend of claiming data as property inhibits scientists' data sharing because scientists view data sharing as losing commercialization opportunities (Tenopir et al. 2011). Generally, scientists believe that formal intellectual property law does not apply to data sharing practices, and their scientific data sharing practices rely more on their own policies, practices, and norms (Fisher et al. 2010). However, scientists in some research disciplines would claim their intellectual property toward their research findings because of commercialization (Tenopir et al. 2011). Concerns about intellectual property are significant in the disciplines where scientists can file patents and potentially commercialize their research (Blumenthal et al. 2006; Taylor 2007). Previous studies showed that the scientists who intend to file patents and monetize their research findings are more likely withhold their data (Blumenthal et al. 1997; Blumenthal et al. 1996).

Scientists' concerns about misuse and criticism of data also decrease the prevalence of data sharing. Scientists fear that their data will be misused or used without appropriate attribution (Borgman 2007; Cragin et al. 2010; Pryor 2009). Sterling and Weinkam (1990) indicated that scientists are reluctant to share their data because other scientists may misinterpret their findings, which may lead to bias or accusations of research fraud. Sedberry and colleagues (2011) indicated that scientists may misuse or misinterpret original data because they lack the original context in which the data were collected.

Vickers (2006) reported that clinical trialists seem to be concerned with misinterpretation of their research. Along with concerns about misuse of data, some scientists also have concerns about potential criticism from other scientists based on possible errors (Liotta et al. 2005). Similarly, Sterling and Weinkam (1990) found that scientists are reluctant to share data because of the potential for conflict and disagreement between scientists.

Lastly, the perceived sensitivity of data also prevents scientists from sharing data. Previous studies show that scientists do not want to share their data because of privacy in the case of human subject research (Lane et al. 2010; 2009) and sensitivity of data for national security (Sterling et al. 1990). Borgman (2009) indicated that data sharing is limited in the fields where human subject research is prevalent, such as social science and biomedical science. Lane and Schur (2010) and Savage and Vickers (2009) found that data sharing can be difficult in health care related fields because of patients' privacy concerns (e.g. HIPAA's privacy rule). More practically, informed consent agreements may not allow scientists to reuse original data in subsequent studies (Piwowar 2010). Previous research has also found that scientists avoid sharing data when they feel the data itself is sensitive (Crall et al. 2010; Sterling et al. 1990). For example, Sterling and Weinkam (1990) found that scientists oppose the international exchange of scientific data due to national security concerns. Crall and colleagues (2010) also found that 27% of the citizen science groups studied were concerned about data sharing because of the sensitivity of the data they collected on endangered species.

Other Individual, Organizational, and Environmental Factors

Previous research also identified other individual, organizational, and environmental factors influencing scientists' data sharing. In regards to individual factors, Blumenthal and colleagues (2006) reported that scientists' prior negative experiences and their mentors' discouragement are significantly associated with verbal or written data withholding among geneticists and other life scientists. In regards to organizational factors, Campbell and Bendavid (2003) reported that according to a survey of 79 technology transfer officers, research universities' institutional policies prevent scientists at those universities from sharing research materials without a material transfer agreement. In regarding to environmental factors, Tenopir and colleagues (2011) found that the decision to share data relies on what stage of publication research is in when others request the data. Other scholars have found that competiveness in either research labs or scientific communities negatively influence scientists' data sharing (Tenopir et al. 2011; Vogeli et al. 2006). In the context of a research laboratory or group, the competition for recognition positively influences data sharing behaviors within that research group or lab (Vogeli et al. 2006), and similarly, in the context of a research community, the competiveness of a field of research negatively influences scientists' data sharing behaviors within that field (Tenopir et al. 2011). Table 2.3 below shows the summary of prior studies on the factors influencing scientists' data sharing:

Data Sharing Factors			Studies
Institutional Factors	Funding agency’s Policy		McCullough et al. 2008; Piwowar et al. 2008a; 2008b; Piwowar et al. 2010
	Journal Requirements		Piwowar et al. 2008b; Piwowar et al. 2010; Noor et al. 2006; Savage et al. 2009
	Contract with Industry Sponsors		Louis et al. 2002; Campbell et al. 1998; Blumenthal et al. 2006; Campbell et al. 2003 (Government)
	Organizational Policies		Campbell & Bendavid 2003;
	Competiveness of Environments		Vogeli et al. 2006 (Labs); Tenopir et al. 2011 (Scientific Communities)
Resources	Metadata standard		Bowker et al. 2000; Zimmerman 2007; Michener 2006; Karasti et al. 2010;
	Data repositories		Marcial et al. 2010; Cragin et al. 2010; Fennema-Notestine 2009;
Individual Factors	Characteristics		Gender (Blumenthal et al. 2006), Prior Experience (Piwowar 2011), Career level (Piwowar et al. 2010)
	Perceived Benefits		Kim 2007; Kling et al. 2003 / Kankanhalli et al. 2005
	Reciprocal Benefit		Zimmerman 2007; Louis et al. 2002 (Internal Sharing)
	Perceived Efforts		Campbell et al. 2002; Louis et al. 2002; Foster & Gibbons 2005; Kim 2007; Tenopir et al. 2011
	Perceived Risks	Losing Publication Opportunities	Reidpath et al. 2001; Savage et al. 2009; Campbell et al. 2002
		Losing Commercialization Opportunities	Tenopir et al. 2011; Blumenthal et al. 2006; Blumenthal et al. 1997; Blumenthal et al. 1996; Taylor 2007
		Misuse	Borgman 2007; Cragin et al. 2010; Pryor 2009; Vickers 2006
		Privacy	Lane et al. 2009; Borgman 2009; Savage & Vickers 2009
		Sensitivity of data	Crall et al. 2010
		Potential Criticism	Liotta et al. 2005

Table 2.3 Summary of Prior Studies on the Factors Influencing Scientists’ Data Sharing

2.3.3. Consequences of Data Sharing

Previous studies in data sharing have studied the benefits of data sharing and the consequences of data withholding. From survey, interviews, and focus groups scholars identified major benefits of data sharing. First, scientists validate previous research by peer review of the original data (Fienberg 1994; Fienberg et al. 1985). By reanalyzing the original data, scientists can confirm or refute research findings (Borgman 2007; Fienberg 1994), which helps prevent scientific error or misbehaviors such as fraud or selective reporting (Vickers 2006). As such, data sharing supports open and transparent scientific research (Borgman 2007; Campbell et al. 2002; Krathwohl 1998). Second, scientists can also test secondary hypotheses using existing data sets (Borgman 2010; Fienberg 1994; Fienberg et al. 1985; Vickers 2006), and can conduct meta analyses (Vickers 2006), which eventually lead to new scientific innovation (Borgman 2010; Campbell et al. 2002; Tenopir et al. 2011). Similarly, scientists can build better research using other scientists' shared data (Vickers 2006). Data sharing allows scientists to advance science by building on other scientists' works (Louis et al. 2002). Lastly, the data shared can also be used to educate science trainees (Vickers 2006). Campbell and colleagues (2002) found that scientists believe that the free and open sharing of publication related information, data, and materials is a critical tool for educating their students.

Throughout the national survey and interviews, researchers identified the consequences of data withholding in their research communities. One of the main consequences of data withholding is that it hinders the scientific research progress (Blumenthal et al. 2006; Vogeli et al. 2006). Campbell and colleagues (2002) reported that data withholding prevents scientists from confirming, replicating, and building on previous published research. The same study also found that geneticists were more likely to report the negative influences of data withholding on their research progress than were other life scientists. Some researchers reported that data withholding also ruined trust and collegiality among researchers (Blumenthal et al. 2006). A more recent Vogeli's (2006) study found that researchers who had denied other's requests or who had their own requests denied reported that data withholding had significant negative influences on the quality of their education, communication in their research group, and their relationships with their colleagues.

2.4. Data Reuse

Relevant issues of data reuse as the extension of data sharing are reviewed in this section. Data sharing is possible based on the premise that the data collected has continuing value for future reuse beyond its original value (Pienta et al. 2010). Uhlir (2010) also argued that the value of data increases when scientists can make more use of the data. The reuse of scientific data can be defined as the secondary use of data collected for one purpose to solve one or more additional research questions (Zimmerman 2008). Data reuse can be understood as active sharing, or the final goal of data sharing. Scientists reuse data for purposes similar to the purposes behind data sharing, such as understanding general trends, confirming or reputing original research findings, providing trainees with educational sources, and encouraging data use in policy making and evaluation (Faniel 2009; Zimmerman 2008).

Previous studies have paid comparatively little attention to the reuse of data (Zimmerman 2008), and very few studies have been done specifically in the area of data reuse (e.g.

Birnholtz et al. 2003; Carlson et al. 2007; Wallis et al. 2006). In the perspective of practices, the data management policies by NIH and NSF do not exactly cover the reuse of data (National Institutes of Health 2003; National Science Foundation 2008).

Previous studies identified various reasons scientists do not actively reuse others' data. First, there is little incentive to use others' data (Sterling et al. 1990). Second, scientists may have difficulty locating necessary data sets because there is no data repository in their scientific communities (Marcial et al. 2010). Scientists need to negotiate data ownership and related issues with the original data producers or the copyright owners (Van House et al. 1998). Finally, with regards to the data itself, shared data often does not contain enough information to be reusable. Data producers do not always consider the extent to which others can use their data (Baker et al. 2007; Cragin et al. 2010).

Various factors can facilitate the reuse of data: improved data repositories and associated infrastructures, complete data, trust among scientists and regarding data, and contextual information (Carlson et al. 2007; Jirotka et al. 2005). Prior studies identified both trust and the context of data as critical factors influencing data reuse (Carlson et al. 2007; Jirotka et al. 2005). Since the data are contextualized where the data originally collected, the researchers need to trust and understand data within the context that it was originally collected in order to properly reuse it (Cragin et al. 2006; Jirotka et al. 2005; Zimmerman 2008).

Trust of data is an important factor influencing data reuse. Trusting data means believing in its quality and provenance (Carlson et al. 2007). Scientists evaluate the reusability of data by assessing its trustworthiness based on their previous experiences (e.g. field knowledge) (Borgman 2007; Faniel et al. 2010; Zimmerman 2007), relevant documentation (Wallis et al. 2007), and their trust in their colleagues (Cragin et al. 2006; Zimmerman 2007). A study of habitat ecologists indicated that scientists may examine any and all documentation related to their colleagues' data collection before they actually feel they can trust and reuse their colleagues' data (Wallis et al. 2007). Cragin and Shankar (2006) and Zimmerman (2007) found that trust among scientists can facilitate scientists' data reuse by increasing the extent to which data are trusted.

Scholars have also considered the limitations of metadata and the necessity of contextual information for actual data reuse. Although metadata can facilitate scientists' data sharing, scholars argued that current metadata models are not enough to support scientists' data reuse (Birnholtz et al. 2003; Bourne 2005; Cragin et al. 2010). Edwards and colleagues (2011) even posited that metadata may cause friction between scientific collaborators and hinder data sharing and reuse. For this reason, scientists treat both specific details and metadata as contextual information necessary to help them comprehend others' original data (Zimmerman 2008).

Therefore, scholars argued that contextual information is critical for data reuse (Birnholtz et al. 2003; Carlson et al. 2007). Bowker and Star (1999) argued that the interpretation of scientific data is an active and context-dependent process, so metadata are insufficient information to provide the data reuser with the full context in which the data were originally collected (Cragin et al. 2010). For this reason Zimmerman (2007) indicated that informal communication between data producers and reusers is often necessary to help scientists to understand the raw data. Contextual information can help scientists reuse data by making the raw data more useful and accessible in complete and accurate details (Baker et al. 2009; Zimmerman 2008). Markus (2001), however, argued that it is very difficult to capture all kinds and sufficient amounts of contextual information necessary to let others reuse data.

2.5. Limitations of Previous Studies

Although previous studies in scientists' data sharing provide valuable insights, they are limited in terms of main focus, research methods, theoretical frameworks used, what research constructs are employed, and what disciplines are studied. First, previous studies have focused mainly on individual motivational factors and technical factors in scientists' data sharing behaviors. However, Tenopir and colleagues (2011) argued that effective data sharing does not just depend on those factors; it is influenced by the practices and culture of everyone involved in the research process as well as by researchers' perceptions. Since scientists' data sharing is influenced by individual motivations, institutional pressures, and facilitating resources, future studies need to consider those factors.

Second, the majority of previous studies did not use any explicit theoretical model to explain scientists' data sharing behaviors. There are not many theoretical models currently exist to guide research on scientists' data sharing. Previous studies have focused on the prevalence of, benefits and consequences of, and factors affecting scientists' data sharing and withholding (Blumenthal et al. 1997; Blumenthal et al. 2006; Campbell et al. 2002; Campbell et al. 2000; Cragin et al. 2010; Kim 2007; Louis et al. 2002; Piwowar 2011). These studies do not employ any explicit theoretical background or identify causal paths among different factors influencing data sharing. Those studies use baseline surveys to understand the percentage of each factor.

Third, previous studies identified few research constructs regarding the factors influencing scientists' data sharing. They found institutional factors (funding agencies and journals' pressures), individual factors (characteristics, rewards, effort, control, fear of misuse, and criticism), and resource factors (metadata and data repositories); however, they focused more on individual perception factors rather than on disciplinary and organizational factors. Additionally, those constructs studied were not synthesized as a research model and were studied sporadically. For example the factors of normative pressure in a research discipline, scholarly altruism, individual attitude, and scientists' self-efficacy toward information management all may influence scientists' data sharing behaviors, but these factors have not yet been studied.

Fourth, previous studies did not cover diverse science and engineering disciplines in regards to scientists' data sharing behaviors. Much of the prior research has focused on life scientists, geneticists, medical researchers, ecologists, and psychologists, rather than on scientists' data sharing behaviors across a variety of science and engineering disciplines. Studies within each discipline also have a limited research scope and extensiveness. As scientists' data sharing varies by discipline (Borgman 2007; Pryor 2009; Tenopir et al. 2011), scientific data sharing behaviors cannot be fully understood without considering disciplinary factors as well as individual motivations. Therefore, more investigation is needed to understand the full picture of data sharing within and between diverse science and engineering disciplines. The multilevel study would be a useful approach to investigate both disciplinary and individual level factors influencing scientists' data sharing behaviors across different disciplines.

Fifth, although previous studies employ a number of research methods to examine the factors influencing scientists' data sharing and reuse, survey was the dominant method used. As such, the current information on scientific data sharing practices is largely limited to data that the survey method can uncover. Scholars indicated that scientists' actual data withholding is more prevalent than what scientists reported in a survey (Blumenthal et al. 2006; Kuo et al. 2008b). Therefore, future research needs to consider qualitative methods or mixed methods to investigate scientists' data sharing behaviors.

By understanding the limitations of previous studies, researchers can develop a theoretical framework to address individual motivations, institutional pressures, and technical resources in research on data sharing. The new theoretical framework would include extensive research constructs including individual, institutional, and resource factors. In addition, this framework would allow researchers to investigate scientists' data sharing behaviors across disciplines rather than focusing on one specific discipline. Lastly, this research framework would employ a variety of data collection methods, including interviews and survey, to provide an extensive picture of scientists' data sharing. This framework can triangulate scientists' data sharing behaviors across different disciplines.

2.6. Summary

In order to understand scientists' data sharing practices, this research considers scientists' norms and values as the structure of science. Scientific norms and values are embedded in scientists' data sharing practices as seen in scholarly communications. Merton (1973) defined the four traditional norms of science as communalism, universalism, disinterestedness, and organized skepticism. Mitroff (1974) provided counter-norms to Merton's four norms of science, including solitariness, particularism, interestedness, and organized dogmatism. Mitroff (1974) argued that Merton's original norms of science and his alternative norms are mixed in an actual science institutions.

Although data sharing is desirable according to scientific communities' norms of communalism and disinterestedness and can contribute to the advancement of scientific research, there is ample evidence that scientists nonetheless withhold their data rather than sharing it in popular science journals (Cohen 1995). Prior studies involving research on diverse factors influencing scientists' data sharing and withholding, can be categorized into three groups, including institutional factors (i.e. funding agency's policy; journal requirements; and contract with industry sponsors); resource factors (i.e. metadata and data repositories); and individual factors (i.e. personal characteristics, perceived benefit, perceived effort, perceived risk).

Although previous studies in scientists' data sharing provide valuable insights, they are limited in terms of main focus, research methods, theoretical frameworks used, what research constructs are employed, and what disciplines are studied. First, previous studies have focused mainly on individual motivational factors and resource factors rather than institutional or disciplinary factors. Second, the majority of previous studies hardly employed any explicit theoretical model to explain scientists' data sharing behaviors. Third, previous studies identified few research constructs regarding the factors influencing scientists' data sharing. Fourth, previous studies did not cover diverse science and engineering disciplines in regards to scientists' data sharing behaviors. Fifth, although previous studies employ a number of research methods to examine the factors influencing scientists' data sharing and reuse, survey was the dominant method used. By understanding the limitations of previous studies, this research discusses possible theoretical frameworks and research methods which can triangulate scientists' data sharing behaviors across different disciplines.

3. Theoretical Framework

This chapter provides theoretical foundations and conceptual model development. Two theoretical perspectives including institutional theory and theory of planned behavior are employed in developing a conceptual model to understand and distinguish both institutional and individual factors influencing scientists' data sharing behaviors. Institutional theory can explain the context in which individual scientists are acting; whereas the theory of planned behavior can explain the underlying motivations behind scientists' data sharing behaviors in an institutional context.

3.1. Institutional Theory

This research employs sociological institutional theory for one of main theoretical foundations. Institutional theory was originally developed to explain organizational behaviors, or why firms adopt similar organizational structures and practices and how they become similar to each other under institutional pressures (DiMaggio et al. 1983). This is called organizational isomorphism, and organizations are hypothesized to be fundamentally influenced by it in order to achieve organizational legitimacy (Deephouse 1996). DiMaggio and Powell (1983) argued that organizational legitimacy can help firms do business with other similar firms by accessing essential resources. Institutional theory emphasizes the way organizations achieve organizational legitimacy rather than productivity or efficiency in an institutional environment (Meyer et al. 1977; Scott 2001).

Institutional theory has evolved over the last several decades, and neo-institutional theory has extended its scope to encompass individuals as well as organizations (Scott 2001). In this study, the term institutional theory mostly means neo-institutional theory developed by modern institutional theory scholars, DiMaggio and Powell (1983), and Scott (2001). Institutional theory can provide significant insights about how social actors are influenced by institutional pressures from their institutional environment. According to institutional theory, social actors face external pressures to conform to shared notions of desirable and appropriate behaviors in order to secure resources and have social support by observing organizational legitimacy (DiMaggio et al. 1983; Tolbert 1985). Social actors not only consider the efficiency or productivity of social behaviors (rationality) but also consider the legitimacy of social behaviors (DiMaggio et al. 1983; Oliver 1991).

Institutions and Institutional Logic

Institutions are considered regulations that constrain individuals' choices and provide predictable conditions (Scott 2001). Institutions can be defined as social structures which include taken-for-granted, formal, or informal rules that restrict social behaviors (Bjorck 2004). Social structures are comprised of symbolic elements, material resources, and social activities (Scott 2001). Scott (2001) defined institutions as "social structures that have attained a high degree of resilience" (p. 48). Institutions are established through institutionalization, which is the process by which rules and behaviors become taken-for-granted and legitimized (Meyer et al. 1977; Tolbert et al. 1983). Once institutions are established, they provide social actors with constraints that work as authoritative guidelines for social behaviors and are taken for granted (DiMaggio et al. 1983; Scott 2004). Individual beliefs form from notions of legitimacy that are constructed by institutions (Barley 1986).

Institutional logic as a shared cognitive framework can be defined as a set of collectively constructed assumptions, beliefs, rules, and practices. Institutional logic provides individuals with principles to help them interpret their experiences and develop their behaviors (Friedland et al. 1991; Haveman et al. 1997; Thornton et al. 1999). Institutional logic, which resides at different levels and fields, is enacted by institutional actors (Chiasson et al. 2005). In the relationship between organizations and individuals, institutional logic on an organizational level ultimately plays out at the level of individual action (Battilana 2006). Thornton and Ocasio (2008) argued that institutional logic shapes individual actions in an organization by providing collective identities that consist of regulative, normative, and cultural-cognitive bases for community members. More specifically, institutional theory scholars also argued that institutional logic shapes people's attitudes and behaviors by structuring incentives (Friedland et al. 1991; Luo 2007).

Institutional pressures

According to institutional theory, an institutional environment provides social expectations and norms, allowing social actors to perform socially-acceptable behaviors, develop socially acceptable practices, and create proper organizational structures and operations (DiMaggio et al. 1983; Meyer et al. 1977; Scott 2001). Social actors need to conform to those social expectations and norms in order to maintain their legitimacy (DiMaggio et al. 1983; Heugens et al. 2009; Zsidisin et al. 2005). Institutional legitimacy as the shared notion of desirable and appropriate actions can be exerted through broader rules, professional norms, and taken-for-granted beliefs (DiMaggio et al. 1983; Meyer et al. 1977; Scott 2001). Scott (2001) identified these pressures as the three pillars of institutions: regulative, normative, and cultural-cognitive. Social actors try to conform to these shared notions of regulative, normative, and cultural-cognitive pressures to achieve and maintain their legitimacy. The details of regulative, normative, and cultural-cognitive pressures are provided below.

The regulative pillar includes coercive aspects of institutions, such as laws or rules, which regulate and constrain actors' behaviors (Scott 2001). The regulative pillar forces compliance through fear of sanctions for disobedience (Scott 2001). Regulative pressures are defined as "both formal and informal pressures exerted on organizations by other organizations upon which they are dependent" (DiMaggio et al. 1983). The regulatory pressure provides individuals with governmental or authoritative power which regulates individuals' behaviors (Scott 2007). Previous studies found that on an organizational level, regulative pressures stem from diverse sources: resource dominant organizations (e.g. suppliers), parent corporations, and regulatory bodies (e.g. government) (Teo et al. 2003). Regulative pressures are sometimes explicitly written as rules and sanctions (Scott 2001).

Normative pressures can be defined as the legitimizing means that stem from collective expectations in a particular institutional context (DiMaggio et al. 1983; Scott 2001). Scott (2001) argued that normative pressures, as collective expectations, are important mechanisms to determine appropriate and legitimate behaviors in a community. Collective expectations become shared norms through training, education, and association (DiMaggio et al. 1983). The main institutions that exert normative pressure include the research community, local networks, affiliations, and certification agencies which espouse public values (Heinrich et al. 2004). Actors are likely to adjust their behaviors according to their beliefs about what other members in the same community view as appropriate (Deephouse 1996).

Cultural-cognitive pressure as a mimetic mechanism occurs “when an organization imitates the actions of other structurally-equivalent organizations that occupy similar economic network positions in the same industry” (Burt 1982). Cultural-cognitive pressures have two main components: the prevalence of a practice in an industry and the perceived success of high-status organizations in an industry (Haveman 1993). Cultural-cognitive pressures push social actors to voluntarily and consciously copy other successful and high-status actors practices and behaviors because they believe those successful actors’ actions are more likely to produce positive results (DiMaggio et al. 1983). Since the cultural-cognitive pillar is rooted in an institutional context, it is difficult to recognize and identify. In other words, the cultural-cognitive pillar is related to a shared understanding of reality that is taken for granted. Actors imitate the practices and behaviors of successful and high-status social actors because they believe that the actions taken by them will be more likely produce more positive results. The three institutional pillars are summarized in Table 3.1:

Component	Regulative	Normative	Cultural-Cognitive
Basis of compliance	Expedience	Social obligation	Taken for grantedness Shared understanding
Basis of order	Regulative rules	Binding expectations	Constitutive schema
Mechanisms	Coercive	Normative	Mimetic
Logic	Instrumentality	Appropriateness	Orthodoxy

Indicators	Rules and Laws Sanctions	Certification Accreditation	Common beliefs Shared logics of action
Basis of legitimacy	Legally sanctioned	Morally governed	Comprehensible Culturally supported

Table 3.1: Scott's Three Pillars of Institutions (Koulikoff-Souviron et al. 2008)

Previous institutional theory based studies have mainly focused on how institutional logic influences organizations and their structures, but less attention has been paid to how institutional logic influences individuals in an institutional environment (Battilana 2006; Vandenabeele 2007; Zucker 1991). Although institutional theory considers that individuals' behaviors are influenced by institutional logic (Scott 2001), previous studies in institutional theory have not systematically investigated how institutional logic shapes individuals' attitudes and behaviors (Rupidara et al. 2011; Szyliowicz et al. 2010). Compared to the macro-level view of institutional theory (DiMaggio et al. 1983), a number of institutional theory scholars argued that institutional theory can be applied to study micro-level phenomena by looking at how institutional pressures influence individuals' beliefs, attitude, and behaviors (Battilana 2006; Hall et al. 1996; Robinson 2011; Robson et al. 1996; Roth et al. 1994; Suddaby 2010; Wicks 2001; Zucker 1977; Zucker et al. 2004).

There are a good number of studies representing micro-level analysis of individual behaviors based on institutional theory (Carney et al. 2009; Kisfalvi et al. 2011; Mezias et al. 1994; Sitkin et al. 2005). For example, Granfield (2007) used institutional theory to identify personality and motivational factors as well as institutional factors that influence lawyers' participation in *pro bono* work. Similarly, scholars used institutional theory to explain individuals' asset building behaviors under a financial program (Johnson et al. 2010; Ssewamala et al. 2004), and they even acknowledge that individual-level theories must be combined with institutional theory (Ssewamala et al. 2004). Some research has even been done on cognitive aspects of institutional theory (George et al. 2006; Powell et al. 2008). Sometimes, neo-institutional theory even considers how individual actors can influence their institutions (e.g. institutional entrepreneurs) (Phillips et al. 2007), and emphasizes the role of actors in shaping institutional processes (Garud et al. 2002; Greenwood et al. 2006; Lam 2010; Oliver 1991).

3.2. Theory of Planned Behavior

This study employs theory of planned behavior as an individual motivation theory, which can be connected with institutional theory. The theory of planned behavior, and its precursor, the theory of reasoned action, are well-established social psychology theories that describe how salient beliefs influence behavioral intentions and subsequent behavior (Ajzen 1991; Fishbein et al. 1975). The theory of planned behavior provides insights regarding how an individual's attitudes, subjective norms, and perceived behavioral controls influence his or her behavior mediated by intention. Along with institutional theory, theory of planned behavior can explain how individual scientists make their decision based on their own motivations. This section reviews both theory of reasoned action and theory of planned behavior as theoretical foundations for individual motivation theory in this research.

Theory of Reasoned Action

Fishbein and Ajzen's (1975) Theory of Reasoned Action (TRA) explains an individual's behavior based on his or her behavioral intention, which is in turn influenced by his/her attitude toward the behavior and perception of subjective norms regarding the behavior. According to Fishbein and Ajzen (1975), behavioral intention refers to "a person's intentions to perform various behaviors," and attitude and subjective norms are defined as "a person's favorable or unfavorable evaluation of an object (or behavior)" and "a person's perception that most people who are important to him/her think he/she should or should not perform the behavior." Attitude and subjective norms are determined by a person's behavioral and normative beliefs (Fishbein et al. 1975). Behavioral beliefs refer to an individual's deeply held opinions and ideas about the consequences of a given behavior, whereas normative beliefs are a person's deeply held opinions and ideas about the perceived expectations of specific referent individuals or groups for his/her behaviors (Fishbein et al. 1975). The theory of reasoned action model is shown in Figure 3.1:

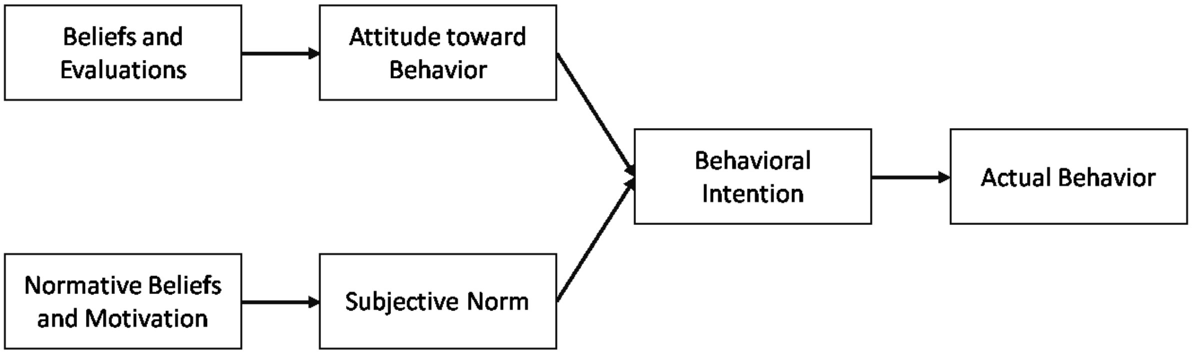


Figure 3.1 Theory of Reasoned Action (Fishbein et al. 1975)

Theory of Planned Behavior

Similar to theory of reasoned action, the Theory of Planned Behavior (TPB) is a well-established social psychology theory also stating that specific salient beliefs influence behavioral intentions and subsequent behavior (Ajzen 1991). Theory of planned behavior added another construct to theory of reasoned action's framework, perceived behavioral control, which means "one's perceptions of his/her ability to act out a given behavior easily" (Ajzen 1991). In TPB, each of the determinants of behavioral intention including attitude, subjective norm, and perceived behavioral control is in turn determined by underlying belief structures including behavioral, normative, and control beliefs (Ajzen 1991).

In the theory of planned behavior, attitude, subject norm, and perceived behavioral control are the key components which explain behavioral intention. In last decades both theory of reasoned action and theory of planned behavior have been applied in diverse social scientific disciplines and have received significant empirical supports. The theory of planned behavior is depicted in Figure 3.2:

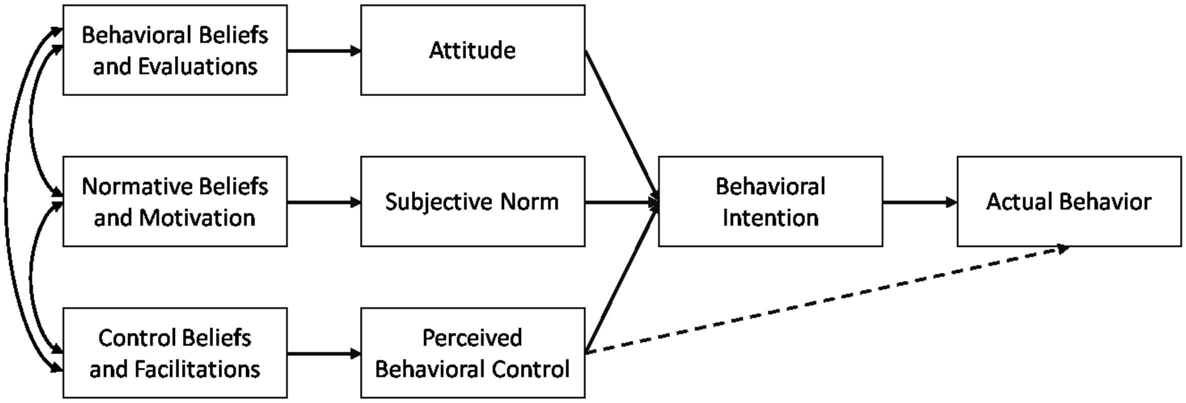

Behavioral Beliefs and Evaluations
Attitude
Normative Beliefs and Motivation
Subjective Norm
Behavioral Intention
Actual Behavior
Control Beliefs and Facilitations
Perceived Behavioral Control

Figure 3.2 Theory of Planned Behavior (Ajzen 1991)

First, attitude toward a particular behavior has been found to predict individuals' intention to perform that behavior (Ajzen et al. 1980; Fishbein et al. 1975). Prior empirical studies support the relationship between attitude and behavioral intention (Hsu et al. 2008; Pavlou et al. 2006; Wu et al. 2007). For example, in technology adoption and use literature, the relationship between attitude and intention has received empirical support (Dickinger et al. 2008; Titah et al. 2009). In knowledge (information) sharing literature, attitude has been examined and found to positively and significantly influence behavioral intention to share knowledge (Bock et al. 2005; Kolekofski Jr et al. 2003). In this research, attitudinal beliefs are considered as important motivational factors influencing scientists' data sharing behaviors.

Second, subjective norms have been studied in different areas of research including technology adoption (Hsu et al. 2004; Venkatesh et al. 2000), knowledge sharing (Kuo et al. 2008a; Kuo et al. 2008b; Ryu et al. 2003), and marketing (Swan et al. 1989). For example, in prior technology adoption studies subjective norm was found to influence individuals' intention to adopt and use technologies (Hsu et al. 2004; Venkatesh et al. 2000). In regards to knowledge sharing, Ryu and colleagues (Ryu et al. 2003) found that subjective norms positively influence physicians' intention to share their knowledge with others through direct and indirect paths. However, in the existing literature on data sharing, researchers have rarely studied how subjective norms influence scientists' data sharing behaviors.

Third, perceived behavioral control refers to people's perceptions of the ease or difficulty of conducting a particular behavior and the amount of control they need to have over the behavior (Ajzen 1991). Perceived behavioral control was introduced to explain situations in which people lack volitional control over their targeted behaviors (Ajzen 1991). Ajzen (1991) argued that if a behavior is not controllable, people are not likely to consider performing it. Perceived behavioral control can be broken down into two smaller constructs: internal behavioral control (self-efficacy) and external behavioral control (resource-facilitating conditions) (Ajzen 2002; Armitage et al. 1999; Manstead et al. 1998).

Internal behavioral control, or self-efficacy, is a construct proposed by Bandura (1986) and is defined as an individual's subjective judgments of his or her capabilities to perform a behavior (Bandura 1986). Compared to self-efficacy, internal perceived behavioral control, which focuses on individual's own capability to perform a behavior, external perceived behavioral control is defined as individual judgments about the availability of facilitating resources and environments to perform a behavior (Ajzen 1991; Hsu et al. 2004; Taylor et al. 1995). In the study of knowledge sharing, scholars found that perceived behavioral control was a significant predictor of intention to share knowledge (Husted et al. 2002). Ryu and colleagues (2003) found that perceived behavioral control influences physicians' intentions to share their knowledge. Kuo and Young (2008b) also found that perceived behavioral control actually precedes the intention to share knowledge. This research considers resource-facilitating conditions to be external behavioral controls at the institutional level.

The limitations of theory of planned behavior and theory of reasoned action are that these theories only consider personal factors rather than any institutional or social factors (Shi et al. 2008). Prior studies employing theory of planned behavior used the de-contextualized model of individual level analyses (Shi et al. 2008). For example, the studies employing external behavioral control (i.e. resource-facilitating conditions) were criticized because they included a non-individual level construct in their theoretical models and tested the models by considering the external behavioral control as the same individual level construct (Hsu et al. 2004). Although theory of planned behavior can explain individuals' motivations and actions, it has its limitations in explaining any contextual factor regarding their behaviors. The theory of planned behavior as an individual level theory does not fully explain scientists' data sharing behavior, so it is necessary to combine it with institutional theory to explain scientists' data sharing behaviors under their institutional contexts. In the next section, I present the conceptual model development based on both institutional theory and theory of planned behavior.

3.3. Conceptual Model Development

Drawing upon institutional theory and the theory of planned behavior, this research proposes a conceptual 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 et al. 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).

Institutional Perspective

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

Returning to Scott's (2001) three pillars, 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). 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.

These three pillars of institutional pressure map onto individual scientists' data sharing behaviors in the context of research communities. Firstly, institutions have regulative pressures that they apply to foster desired behaviors. As resource-dominant organizations, the funding agencies that support scientists' research may create regulative pressures for scientists to share data as a condition of their funding. Also, journal publishers exert regulative pressures on the authors of scientific articles through editorial policies on data sharing. Secondly, scientific disciplines and professions may have their own social expectations that encourage or discourage data sharing. Social expectations based on shared norms in scientific communities provide scientists in those communities with normative pressures to share data. Scientific communities may have collective expectations about data sharing based on shared norms (e.g. communalism), and these collective expectations pressure scientists to share their data. In effect, as institutional and disciplinary pressures on data sharing increase due to increased data sharing among colleagues within a scientific community, individual researchers respond to these pressures with some consideration of the merits of participating in the trend (Scott 2001; Tolbert et al. 1983). Lastly, scientists may take data sharing for granted as a part of their culture in their scientific communities. A shared understanding of data sharing in a scientific community provides cultural cognitive pressures for scientists to imitate approved practices and behaviors without individual cognitive processes. In this case, data sharing is deeply embedded in research communities as constitutive schema (Scott 2001).

Traditional institutional theory has focused on how regulative, normative, and cultural-cognitive pressures legitimize organizational structures and practices in a given sector, and on how this legitimacy tends to foster organizational isomorphism across organizations within the sector. However, this research is more concerned with how these pressures influence individuals' behaviors in an institutional context. Scott's (2001) neo-institutional theory can explain how the three pillars of institutions influence scientists' data sharing behaviors at an individual level from the perspective of legitimacy and isomorphism. Individual scientists seek legitimacy through data sharing under institutional pressures, but individual scientists also behave based on individual motivations stemming from their own beliefs and perceptions. Along with institutional theory, the theory of planned behavior can help to explain individual scientists' data sharing behaviors based on their own motivations from perceptions.

Individual Perspective

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 et al. 1975). Theory of planned behavior explains an individual's behavior based on his or her behavioral intention, which is 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 et al. 1975).

Using the perspective from the theory of planned behavior, scientists' data sharing behaviors can be explained by behavioral intentions emerging from: (1) the attitudes they form from their behavioral beliefs and evaluations of the "outcomes" of data sharing; (2) their understanding of subjective norms around data sharing coming from "close colleagues" expectations; and (3) the perceived controllability of their data sharing behaviors.

First, scientists' attitudes toward data sharing influence their intentions to share data. Scientists' behavioral beliefs and their evaluations of the consequences of data sharing lead them to form attitudes toward data sharing. Second, subjective norms influence scientists' data sharing intentions. The subjective norm in the theory of planned behavior is a concept similar to that of normative pressures in institutional theory. In contrast to normative pressure, which comes from virtually connected other scientists in their fields (Meyer et al. 1977; Scott 2001), subjective norms come from "close colleagues" in their interpersonal social network. Lastly, Perceived Behavioral Control (PBC) may influence scientists' data sharing behavior. Scientists can form their perceived behavioral controls from both internal PBC and external PBC. Internal PBC is similar to the construct proposed by Bandura (1986) – self efficacy – that reflects judgments of one's own capabilities to enact a behavior successfully. With respect to data sharing behavior, a sense of internal PBC may arise from scientists' expertise (or lack thereof) in using the tools and technologies that facilitate data sharing. External PBC is an individual judgment about the availability of resources and opportunities to perform the behavior

(Hsu et al. 2004). A researcher's judgments about the availability of IT support within a team or organization, and the existence of data sharing protocols, procedures, and data repositories, may influence how likely they are to engage in data sharing (Hsu et al. 2004).

Underlying Assumptions

This study combines institutional theory and the theory of planned behavior. In order to integrate two different theories, it is important to understand their underlying assumptions. The main assumption behind the theory of planned behavior is that individuals are rational and make reasonable decisions based on their attitudes, subjective norms, and perceived behavioral controls (Ajzen 1991; Fishbein et al. 1975). Although the theory of planned behavior assumes individuals' rationality, it does not imply that all behaviors are necessarily rational from an objective point of view (Contento 2011). The core assumption of institutional theory is that social actors respond to institutional influences to conform (DiMaggio et al. 1983; Scott 1995). Institutional theory basically rejects the assumption of rational choice theory that social actors are rationally seeking to maximize efficiency and productivity (DiMaggio et al. 1983; Scott 1995). In other words, institutional theory assumes that individual actors do not conduct their behavior based on 'pure' rationality; they pursue acceptable performance to legitimize their behaviors along with rationality in an institutional context (Budros 2002). Therefore, the integration of institutional theory with the theory of planned behavior can provide a complementary view of scientists' data sharing behaviors by focusing on the conformity to legitimacy and individual motivations of behavior together.

Previous studies have already combined both institutional theory and individual-level theories to understand individuals' behaviors. For example, Shi, Shambare, and Wang (2008) connected institutional theory and the theory of reasoned action (Ajzen 1991; Fishbein 1980; Fishbein et al. 1975) to examine the adoption of Internet banking. Similarly, Teo, Wei, and Benbasat (2003) and Son and Benbasat (2007) used institutional theory to examine top executives' and high-level managers' intentions to adopt inter-organizational systems and, they brought the concept of intention from Ajzen and Fishbein (1980)'s work.

The conceptual model (Figure 3.3) below provides an extensive map of scientists' data sharing behaviors and shows how scientists make their own decisions to share data based on both institutional theory and theory of planned behavior. In addition, this conceptual model considers institutional resources as important underlying infrastructures supporting scientists' data sharing behaviors.

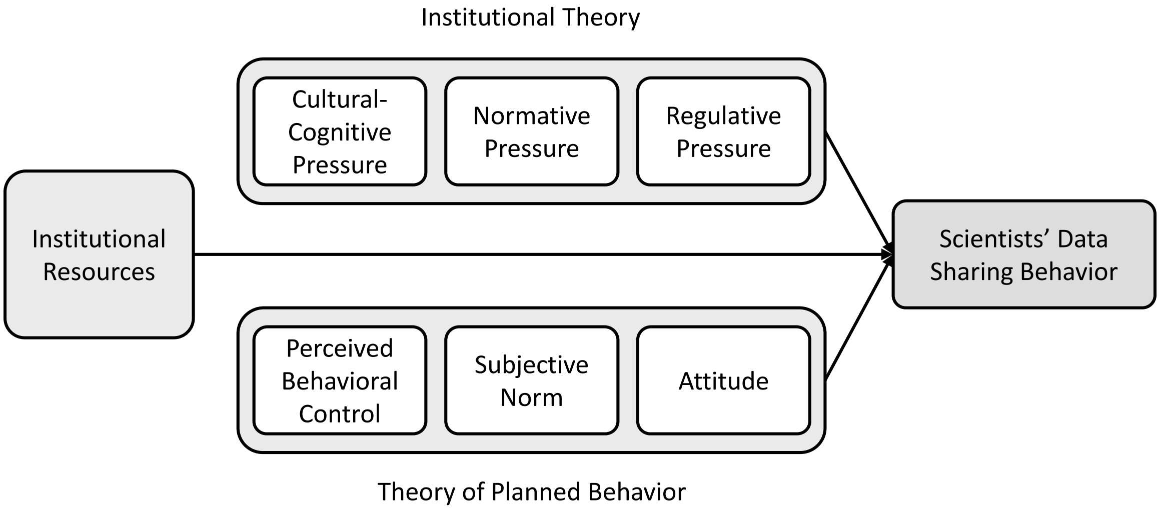

Institutional Theory
Cultural-Cognitive Pressure
Normative Pressure
Regulative Pressure
Institutional Resources
Scientists' Data Sharing Behavior
Perceived Behavioral Control
Subjective Norm
Attitude
Theory of Planned Behavior

Figure 3.3 Conceptual Model for Scientists' Data Sharing Behaviors

3.4. Summary

Drawing upon institutional theory and the theory of planned behavior, this research proposes a conceptual model to investigate how both institutional and individual drivers influence scientists' data sharing behaviors. Scientists' data sharing behavior can be understood through the lens of individual motivation and institutions' seeking organizational legitimacy. Institutional theory (Scott 2001) provides significant insights regarding the importance of institutional environments including organizational rules, norms, and culture on individuals' actions (behaviors) (Tolbert 1985; Tolbert et al. 1983). In contrast, the theory of planned behavior provides its insights regarding how individuals' beliefs influence individuals' behaviors.

This research's conceptual model builds on insights from Scott's (2001) neo-institutional theory. Neo-institutional theory posits three kinds of institutional pressures influencing behaviors: regulative, normative, and cultural-cognitive. Regulative pressure provides organizations or individuals with coercive constraints, and legally sanctions those who do not comply. Normative pressure sets shared norms for the appropriateness of individuals' or organizations' behaviors. The cultural-cognitive institution is deeply embedded in communities and is supported culturally. These three pillars of institutional pressure map onto individual scientists' data sharing behaviors in the context of research communities.

The conceptual model also employs Ajzen's (1991) theory of planned behavior as an individual motivation theory, which can be connected with institutional theory. The theory of planned behavior provides insights regarding how an individual's attitudes, subjective norms, and perceived behavioral controls influence his or her behavior mediated by intention. Along with institutional theory, theory of planned behavior can explain how individual scientists make their decision based on their own motivations. The conceptual model provides an extensive map of scientists' data sharing behaviors based on the combination of institutional pressures and individual motivations.

4. Preliminary Study and Results

This chapter covers the overall research design of this dissertation and the preliminary interview study performed prior to the main survey study. A total of 25 individual interviews were conducted to understand scientists' current data sharing practices. The main purpose of the preliminary study was to explore the landscape of scientists' data sharing practices in difference scientific communities. Results showed support for an institutional perspective on data sharing, as well as an individual perspective for better understanding of scientists' data sharing behaviors. The results of this preliminary study were used to assist in the development of research model and the design of survey.

4.1. Research Design

This research uses a mixed-method approach by combining qualitative and quantitative methods to gain better insight in studying scientists' data sharing behaviors. The exploration of research questions occurred through two interconnected investigations: (1) interviews with scientists in diverse scientific disciplines to understand the extent to which they share their data with other researchers and exploration of institutional and individual factors affecting their data sharing behaviors; and (2) survey research to examine to what extent those institutional and individual factors influence scientists' data sharing behaviors in diverse science disciplines. The overall research procedures with interview and survey studies are presented in Figure 4.1 below.

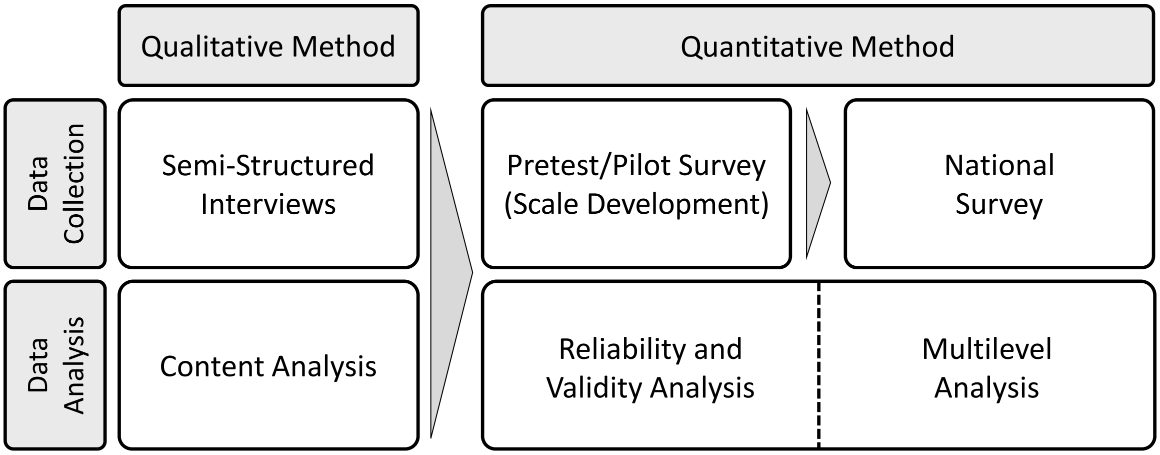


Figure 4.1 Overall Research Procedures with Qualitative and Quantitative Methods

In the first phase, a preliminary study was conducted based on interviews with individual scientists from different disciplines. The main purpose of the preliminary study was to explore the landscape of scientists' data sharing in different scientific communities, as well as the factors influencing scientists' data sharing behaviors. The results of the preliminary study were used to develop a research model for variations in data sharing through the lens of theories that account for individual choices within institutional contexts. Benbasat and colleagues (1987) pointed out that qualitative approaches are suitable for investigating a phenomenon in which research and theory are at their early or formative stages. The results of this preliminary study were used to assist in the development of research model and the design of survey. The detailed research method and analysis for this preliminary study is reported in Chapter 3.

At the second phase, the research model developed at the first stage was tested with a survey method. This research employs a survey as a main research method. Survey is a well-known quantitative research method based on the responses to questions by a sample of individuals in a large population (Punch 2005). The survey method in this research helps to examine the constructs and hypothesized relationships of the scientists' data sharing model. By conducting the survey in diverse science and engineering disciplines, this research can validate the scientists' data sharing model by investigating both institutional and individual influences of scientists' data sharing behaviors.

The preliminary study has limitations in confirming and validating the relationships between the predictors and data sharing behaviors, since it only employed 25 interviews from a limited number of academic institutions in the central New York. The survey method can produce more generalized results about the institutional and individual factors influencing scientists' data sharing behaviors, since surveys employ a probability sampling from the large population (Schutt 2006). The rest of this chapter covers the details of survey method, including population and sampling, instrument development, and reliability and validity issues. Also, the data collection procedure and data analysis plan for the field survey is presented at the end of this chapter.

4.2. Data Collection

From October 2011 to December 2011, I conducted a total of 25 individual interviews to understand STEM (Science, Technology, Engineering, and Mathematics) researchers' current data sharing practices. The main focus of the interviews was two-fold: (1) to explore domain specific data sharing practices in diverse disciplines; and (2) to investigate the factors motivating and discouraging STEM researchers' current data sharing. The Institutional Review Board (IRB) at Syracuse University provided approval of a plan to conduct the individual interviews within three research universities in the eastern U.S. I sent a recruiting email message directly to the STEM researchers, and I also contacted department chairs to distribute the recruiting email message to their STEM researchers. I received 28 responses in total from STEM researchers in three research universities, and I ultimately interviewed 25 interviewees. The remaining three respondents could not be scheduled in time to complete data collection. In order to understand the domain specific data sharing practices in diverse disciplines, I tried to include at least one or two researchers in each research discipline (see Table 4.1).

All the interview sessions were audio-recorded and subsequently transcribed. All the interviews were conducted in English except one interview, which was conducted in Korean for the convenience of the interviewee. I transcribed the interview in Korean and then translated into English for the data analysis. Each interview took 25-35 minutes. I used an open-ended semi-structured interview method by asking similar structured interview questions to all the interviewees including STEM researchers' current data sharing methods, types of data generated and shared, their perceived motivations and barriers of data sharing, and lastly interviewees' demographic information and work environments. An example of the interview questions was: "What motivates researchers (including you) in your field to share their data?" (The preliminary study's interview questions are provided in the Appendix 8.1.) During the interviews, the participants were asked to answer the questions based on not only their own experience but also their observations in their research disciplines in general.

The 25 participants for the interviews include 11 tenured (full and associate) professors, eight assistant professors, one emeritus professor, one professor of practice, two post-doctoral research associates, and two doctoral candidates from three major research universities in the eastern U.S. (17 men and 8 women). Given the goals of this research, I mainly interviewed professors rather than graduate students, but the two post-docs and two senior doctoral students provided perspectives that seemed complementary to the other data, so I retained them in the corpus. The research disciplines of the 25 interview participants are shown in Table 4.1. There were a few minor differences between the names of the departments the interviewees belonged to versus their disciplinary affiliations.

Discipline	Number of Interviewees
Biology	2
Chemistry	3
Computer Science	2
Ecology	5
Electrical Engineering	1
Environmental Engineering	4
Mathematics	1
Mechanical Engineering	2
Physics	3
Radiation Oncology	1
Science Education	1
Total	25

Table 4.1 Research Disciplines of Interviewees

4.3. Data Analysis

The content analysis technique was used to interpret the qualitative data of preliminary interviews. The transcribed interviews were imported into “QDA Miner,” a qualitative data analysis tool. The coding scheme was developed by using both deductive and inductive approaches. I started with ideas arising from neo-institutional theory and individual motivation perspectives to create the data analysis coding scheme. The basic coding scheme included institutional theory-based constructs (regulative, normative, cultural-cognitive pressures); individual motivation-based constructs (benefits, risks, and efforts); and resource constructs (organizational and institutional resources). As I processed the data, I also used an inductive approach to create more specific codes (e.g. scholarly altruism). The interview corpus contained 837 utterances overall; I applied codes to 276 of these utterances regarding the factors both motivating and preventing researchers' data sharing (Table 4.2 only reports the number of respondents out of 25 interviewees in each code; there was only 209 responses in total except 67 redundant responses.).

4.4. Results

The codes revealed STEM researchers' work environments, the types of data they commonly generated, current data sharing methods, and their motivations for and barriers to data sharing. In the following sections, I report on each of these topics by providing a holistic overview of what the codes and their underlying utterances revealed. The coding scheme I used for the motivating and impeding factors of data sharing, a brief explanation of each code, and the numbers of respondents out of 25 interview participants in each code are shown in Table 4.2. The frequency of the factors influencing scientists' data sharing in the form of a radar plot is displayed in Figure 4.2.

Category	Code Name	Brief Explanations	Number of Responses
Regulative Pressures	Funding agency pressure	Funding agencies (e.g. NSF and NIH) require researchers to share their data	16
	Journal's requirement	Journal publishers require researchers to publish their data before their articles are published	9
	Special funding restrictions	Sharing private companies' and military data is restricted	6
Normative Pressures	Professionalism in the fields	Data sharing is a part of their professional mission to develop science	13
	Colleagues' expectations	Feel social pressures by colleagues (being expected to share their data)	7
Cultural-Cognitive Pressure	Colleagues' performance	Observed other colleagues who use shared data and improve their research performance	3
Perceived Benefits	Demonstration of quality work	Shared data indicates the quality of your work; improve the overall research quality	6
	Credits and reputation	Expect credits (e.g. authorship, citations, acknowledgements), reputation, and recognition	15
	Research performance	Conduct a comparative study or large-scale study (novel scientific finding); save time and effort in replicating and collecting data	14
Perceived Efforts	Data annotation	Need to annotate data with their own metadata schemes (no standardized metadata scheme)	10
	Data organization	Takes time to organize data for more understandable, compatible, interoperable formats	11
	Data set location and interpretation	Takes time to find appropriate data sets and understand the data exactly	4
	Technical problems	Being involved with compatibility and interoperability issues with data	9

Category	Code Name	Brief Explanations	Number of Responses
Perceived Risks	Losing publication opportunities	Have less opportunities for future publications; make more exclusive publications if data are not shared	15
	Getting Scooped	Worried about data theft; cannot trust others	8
	Misinterpretation and scrutiny	Worried about having different results by not being analyzed properly or being criticized by others because data are not reliable or low quality	13
Altruism	Altruistic motivation	Allow other researchers to find something interesting that the first people missed; contribute to scientific developments; help others to save time and effort	12
Self-Efficacy	IM/IT expertise	Have technology expertise to manage data	5
Institutional Resources	IM/IT support	Have internal IT/IM supports from their organizations	11
	Data repository	Have data repositories or enough space to share data	9
	Metadata standard	Have data sharing standards (metadata schemes) and systematic procedures	13

Table 4.2 Content Code Explanations and Counts

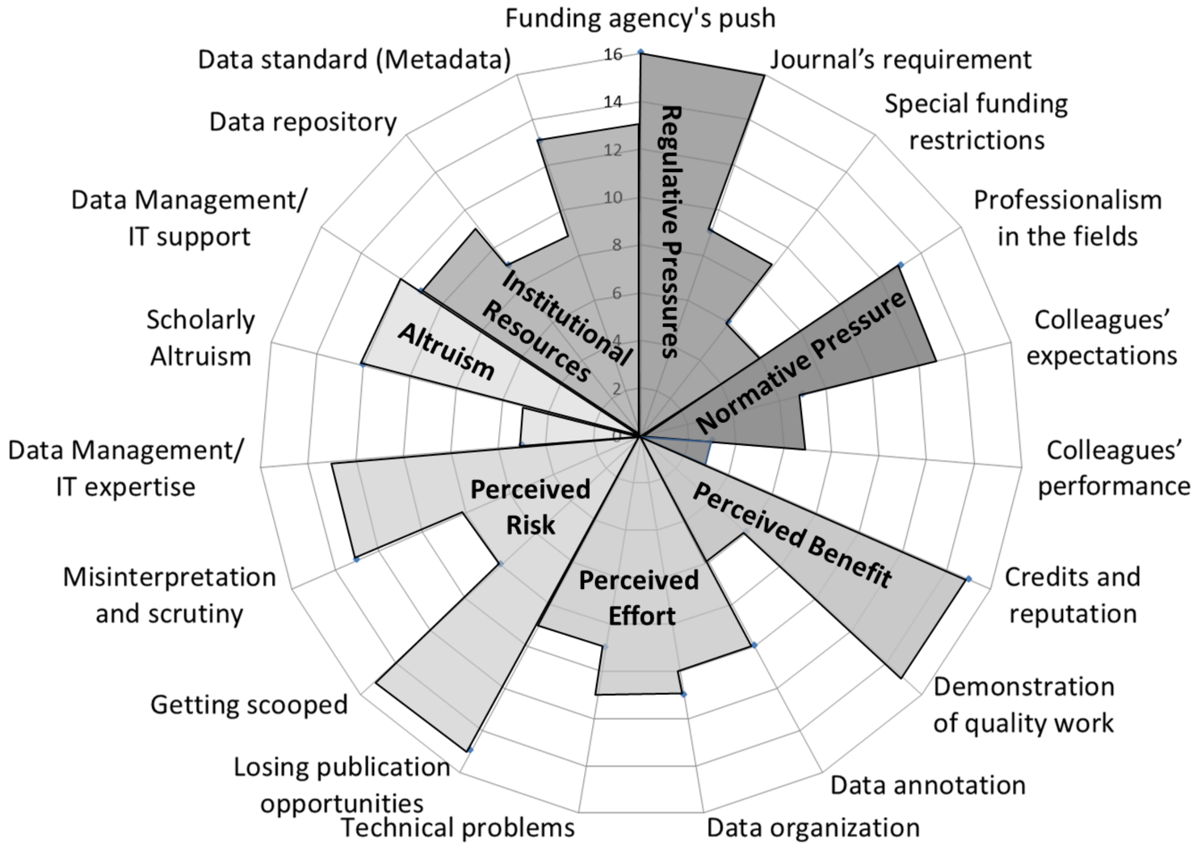


Figure 4.2 Frequencies of the Factors Influencing Scientists' Data Sharing

4.4.1. Research Environment and Data Generated

Most of the interview participants worked in team-based research environments or a mixture of team-based and individual work; only two scholars, a mathematician and theoretical physician mainly worked as individuals. The research teams usually included a lead professor, one or two post-docs, and a few doctoral and masters' students.

The researchers reported that they generated a large amount of domain-specific original data including experimental data (e.g. genome sequencing data, compound data), field data (e.g. soil measurement, animal behavior, tree counts), and computational data (e.g.

software code, computer simulation data). Most of the interviewees felt that they had limited *individual authority* to share their data by acknowledging that sometimes they need to seek permission from others for any collaboratively collected data. Only two interviewees (one post-doc and one doctoral candidate) felt they had *no authority* over sharing the data they collected.

Researchers reported different perceptions of the *importance* of data sharing in their fields. The researchers in biology, chemistry, and ecology agreed that data sharing is critical for novel scientific findings, but the researchers in computer science, electrical engineering, mechanical engineering, mathematics, and radiation oncology disagreed with this belief. Researchers in environmental engineering and physics reported a mixture of both perspectives.

4.4.2. Data Sharing Methods

Researchers in different disciplines reported different data sharing methods. Most researchers reported *internal* data sharing within their research teams or among collaborators; they usually used email, FTP servers, and website as the major internal data sharing methods. I assumed from the start that this type of internal sharing was occurring, and did not investigate further beliefs or motivations in this area.

Researchers also reported diverse forms of *external* data sharing with the researchers outside their research team or collaborators. First, researchers asserted that they share their data upon request; they use email or website upload as method of fulfilling such requests. Researchers also reported contacting other researchers individually to gain access to their data sets from published articles. Across different disciplines, this data sharing method was common, and it was the only data sharing method in the disciplines which do not have any informal or formal data repositories.

Second, some researchers who do not have any formal data repositories in their disciplines used a personal website to share their data with other researchers. A group of scholars in a similar research subject develop an informal or *ad hoc* data repository and share data with other researchers in the research subject area.

Third, some disciplines including biology, chemistry, and ecology use a range of external repositories (e.g. Dryad), and domain-specific data repositories (e.g. GenBank, Protein Data Bank, Computational Chemistry Database, Crystallography Open Database, Long Term Ecological Research Data Repository). These researchers reported well-developed data sharing protocols including data repository and data and metadata standards. In these same disciplines, most of the journals require researchers to publish their data in data repositories.

Finally, researchers in certain disciplines such as chemistry – where there are small, but highly structured data sets – share their data as an electronic supplement through the journals' websites. For example, some scholars in chemistry share their compound data through their journals' online supplements.

Some researchers reported an explicit expectation of various types of professional credits for data sharing including co-authorship, citation, and acknowledgement when their shared data are used by other researchers. There was insufficient data to judge the differences for these expectations among different disciplines, but I noted that the researchers whose disciplines have well established data sharing practices expected less credit than the researchers who do not have any formal way of data sharing. Additionally, I noted that junior researchers had higher expectations for credit than senior researchers and mentioned strengthening the tenure case as the primary motivation for this.

Roughly one third of the interviewees reported that researchers in their field generally share their data after publication. The researchers in the disciplines which do not have any formal data sharing mechanism almost always share their data only after publication. For example, researchers in the engineering fields reported sharing their data only after publication. Another third of the interviewees reported that they shared their data right after their data collection or after a fixed embargo period, regardless of publication status. For example, some researchers in biology and ecology shared their data to a data repository right after data collection. These particular researchers reported a strong sense of trust that their colleagues would not “scoop” them using the shared data.

Lastly, where data sharing was a journal requirement, researchers in chemistry and biology and some researchers in ecology shared their data along with their publications. As noted above, these were cases where journals support a simultaneous publication of relatively small, structured data sets as supplements.

In terms of types of data shared, the researchers in some disciplines (e.g. biology, ecology, environmental engineering) shared raw data, but the researchers in other disciplines (e.g. chemistry, physics) share more refined or processed data. Also, the researchers in computer science, computational chemistry, and physics were prone to share both software and simulation results.

4.4.3. Factors Influencing Data Sharing

The primary focus of this research was on the factors influencing researchers' current data sharing practice. Based on the coding I did, I confirmed specific factors both motivating and preventing researchers' data sharing. In the material below, I explain these factors in three separate groups including institutional, individual, and resource factors.

Institutional Factors

Pressures by funding agencies, journal publishers, and private funding organizations influenced researchers' data sharing practice. First, the single most significant motivation for scientists' data sharing (giving) is a push by funding agencies to make data from funded projects available. Scientific funding agencies in the U.S. including National Science Foundation (NSF) and National Institutes of Health (NIH) require their awardees to share the research data from projects they fund. Second, journals' requirement of data sharing is another factor. The journals in biology, chemistry, and some in ecology require their researchers to publish their data in any types of data repositories. Third, private and certain government funding agencies restrict researchers' data sharing. For example, some pharmaceutical companies and military agencies typically do not allow their awardees to share their data.

Disciplinary influences also affected researchers' data sharing. In many disciplines, data sharing is considered part of the professional responsibility; researchers believe that data sharing is one of their missions, and that it will help the development of their research disciplines. In these same disciplines, researchers reported that they are *expected* to share their data; they feel pressure from their colleagues to do so. Researchers reported observing what other researchers do, and they indicated that they tried to follow colleagues' practices that they saw as useful. A few researchers reported a belief that the research performance of other researchers who use the shared data would improve.

Individual Motivation Factors

Researchers also gave evidence that they carefully examined pros and cons of data sharing before they committed to sharing data. First of all, some researchers reported a belief that data sharing could highlight the quality of their work in research. For some, data sharing provided professional "credit" including co-authorship, citation, and acknowledgement, and reputation. In terms of using the shared data, researchers also believed that data sharing would improve their research (e.g. time saving in collecting the same data, replicating data for another research, conducting diverse comparison studies and large scale research).

Researchers also believed that data sharing imposes efforts for them. In some scientific disciplines (e.g. ecology and environmental engineering) researchers saw the importance of data sharing, but they saw data sharing as very costly in time and effort. Due to a lack of established metadata standards and data preparation procedures, they saw the processes of organizing and annotating their data as very expensive. These same researchers also reported technical problems in the data sharing such as data compatibility and interoperability issues. This was a similar finding across each discipline that did not have well-established data sharing standards (metadata), procedures, and repositories. Researchers in those disciplines also reported that it took substantial time to locate and understand other researchers' data since the data do not have any established data repositories and standardized metadata.

Certain perceived risks by researchers also discouraged them from sharing their data with other researchers. Many researchers worried about losing publication opportunities by sharing their data. It took a lot of time and effort to collect data, and they desired having as many publications as possible from their data. These researchers also worried about getting scooped on innovative findings when they shared their data with other researchers. Two scholars in environmental engineering mentioned that "data sharing is a little bit of a threat to our science because it is less incentive to collect your own data when all data are freely shared." Additionally, several researchers considered that misinterpretation and heightened scrutiny of their data would be possible risks if they shared their data.

Altruism emerged in about half of the interviews as a factor influencing researchers' data sharing. Some researchers reported a strong desire to help their colleagues to save time in collecting data and to avoid replicating experiments unnecessarily. Additionally, these researchers believed that their colleagues could exploit the data in ways that would extend the original findings and thereby benefit the scientific area where they collectively worked. These researchers reported a sense of personal satisfaction coming from sharing their data. A couple of the interviewees mentioned the importance of data sharing cross disciplines not only within a discipline. A biologist mention that "it is also critical to improve [data] sharing across disciplines because a lot of research now days is becoming more multi-disciplinary so for example you have engineers working with biologists or physicists working with engineers and especially in my field in tissue engineering its very multidisciplinary field … If scholars in different disciplines could share that information, then the field of tissue engineering would progress a lot faster."

Institutional Resource Factors

Institutional resources were found to be important factors influencing scientists' data sharing practices. I focused my questioning on two distinct areas: an individual's organizational resource to support the relevant IT tools (internal resources), and the availability of appropriate community tools and infrastructure (external resources). Internal resources included any information management and/or IT support from within their own research team or host organization. Researchers with strong internal support in these areas also reported more extensive data sharing and reuse.

External resources referred to supports for researchers to share their data provided by the research community at large. In this area, researchers reported data repositories, metadata standards, and established data sharing procedures as key features. Biologists and chemists reported that they could easily share their data because they have well-developed data repositories, metadata standards, and procedures to share their data with other researchers. Researchers in engineering fields generally did not report any central or domain data repositories. These engineers also reported needing to spend a lot of time to annotate, organize, upload, and manage their data on subject-specific or *ad hoc* data repositories. Researchers in ecology reported that they are aware of the importance of data repositories and metadata standards and they have developed domain specific repositories and subject specific repositories. Since their data were unstructured, however, they reported that they still needed to develop better metadata standards and data sharing procedures.

4.4.4. Changes in Data Sharing

Our interviewees reported that during recent years they had observed changes in their data sharing practices. Many of the interviewees reported that researchers' awareness, funding agencies' push, journals' requirements, technological improvements, and increased availability of data repository as changes they had experienced within recent memory. Just a few mentioned the emergence of data sharing standards as another recent change.

4.4.5. Supports Needed for Data Sharing

I asked the interviewees what kinds of additional supports they needed to facilitate data sharing. Ten of the 25 interviewees mentioned they do not need any supports since they are satisfied with their current data sharing practices. One biologist and one chemist said that they can easily share their data because they have well-established metadata standards, data sharing procedures, and data repositories. However, the remainder of the interviewees mentioned that metadata standards and data repositories are the main concerns of their current data sharing practice. Additionally, two researchers mentioned that they desired a data portal site where they could search available data sets. Several interviewees indicated that they needed better technology support. In particular, they reported that they needed professionals who could manage data sets, databases, storage, and other IT infrastructure.

4.5. Discussion

In this section, I provide the synthesis of my preliminary study's findings. The institutional perspective seems helpful in understanding the preliminary interview data. In the disciplines of biology and chemistry as well as within some areas of physics, researchers seem to have well-established data sharing methods covering the data lifecycle. These methods are supported by many if not all of the institutions in which they are embedded, mainly through the availability of data sharing standards and repositories.

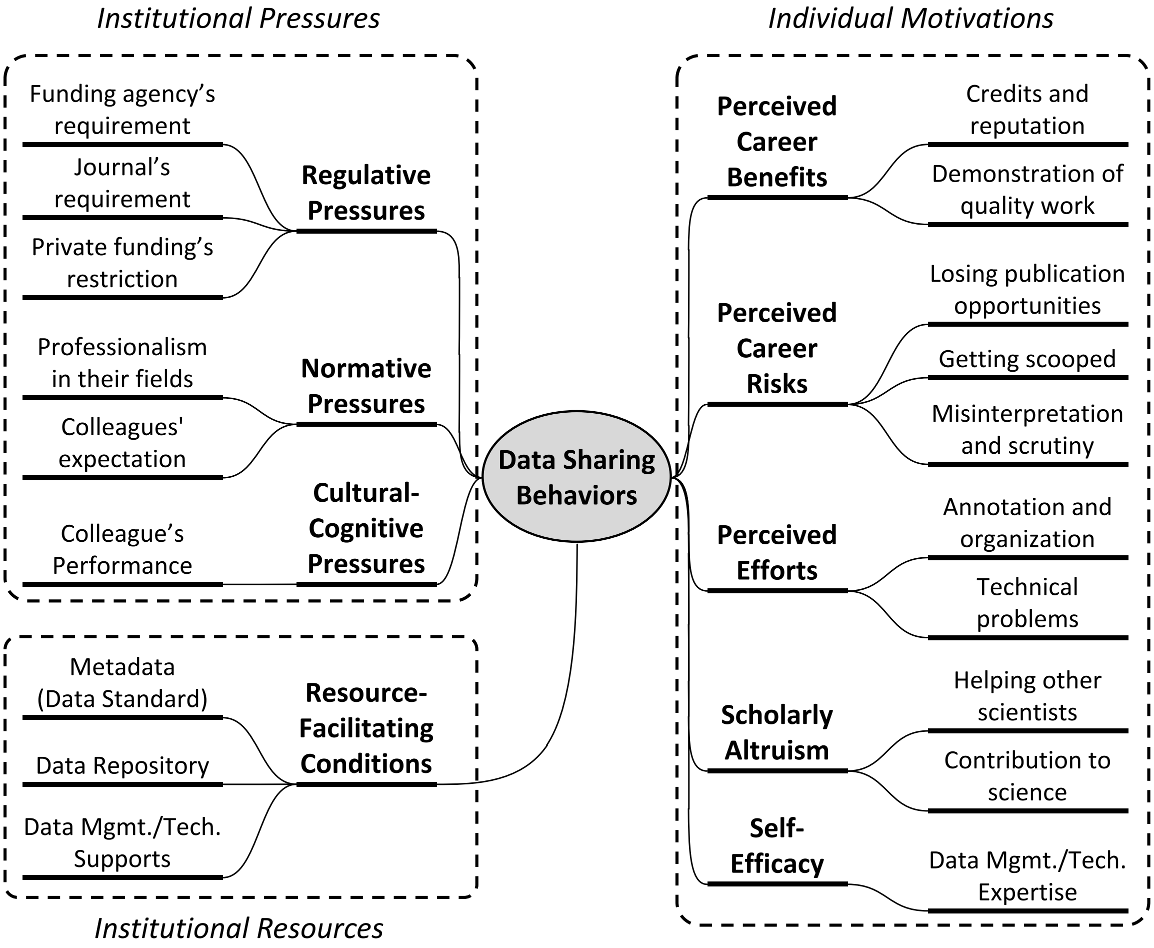


Figure 4.3 Factors Influencing STEM Researchers' Data Sharing Practices

Neo-institutional theory and theory of planned behavior provided a productive lens for reviewing the interview data. Some newer forms of institutional theory incorporate a cross-level perspective by linking institutional forces together with the motivations and behaviors of individual actors. I began this study by framing the situation of the researcher as an individual actor embedded within his or her discipline as well as within the host institution and a variety of external institutions (e.g., funding agencies). Regulative, normative, and cultural-cognitive forces acting on institutions may trickle down to influence the decisions and behaviors of individuals who work within those institutions. An overview of the preliminary findings is provided in Figure 4.3.

To have well-established data sharing practices, researchers need to have supportive institutional environments (e.g. data sharing structures, norms, policies), sufficient resources (e.g. metadata standards, repositories), and positive attitudes toward data sharing (e.g., perceived benefits, efforts, risks). The combination of these can lead to more proactive data sharing practices among researchers. In addition, one surprising finding arose from the spontaneous reports of altruistic motivations for sharing data.

Contrasting biology or chemistry with the discipline of ecology, many ecologists realize that data sharing is critical for their research, but they have difficulties in data sharing because they do not have well-established metadata standards and domain-specific data repositories. For those who do share data, this means spending more time and effort to annotate and organize their data with their own metadata and format. Relatedly, because they do not have well-established central or domain specific data repositories, they share their data through *ad hoc* mechanisms such as Web servers and email exchanges among their collaborative group members. One ecologist mentioned that "[they] should have the official protocol for [data they collected] … those should be peer reviewed and approved and archived just like our data documentation … [they need to] share the procedures not the data only." Researchers also mentioned the importance of having access to information professionals who can support their data sharing in terms of information and technology management. The information professional can help not only share their data, but also use other researchers' data by locating and interpreting the data.

In addition, it seems important to have a central data search mechanism so that researchers can find appropriate data sets for their research. Some researchers mentioned that they have difficulties in locating and interpreting other researchers' data, and they mentioned the necessity of a central data search mechanism. Even in areas where researchers are very good at sharing their data with other researchers, many researchers still do not actively seek other researchers' data sets. Data sharing is a two-way process of providing their own data and using other researchers' data. In order to achieve the promise of data sharing, researchers need to not only provide their data, but also use other researchers' data more actively.

Finally, and perhaps most importantly, this study indicated the importance of aligning institutional pressures with individual motivations for professional achievement. The most frequently mentioned driver of data sharing behavior was the "push" by the funding agencies that support research to ensure that data from the projects they support are made available to other researchers. This force, together with pressure exerted from scholarly journals, can have a strong influence over time on the choices and activities of individual researchers. Ultimately, the advocacy of funders and journals will also need to reflect on universities' policies and mechanisms for promotion and tenure in order to have a more direct influence on the data sharing activities of researchers. When sharing (and reuse) of data leads directly to an improvement of professional reputation and resulting career rewards, researchers will have strong individual motivations to participate in data sharing and reuse.

Taken together, the results support the idea that when institutional forces, infrastructure, and individual motives converge, the behavior of individual researchers will change in response. Many of the researchers I interviewed reported having seen this convergence and these changes during the course of their own careers. Further research efforts are needed to examine the role that altruistic motivations may play in establishing a virtuous cycle of data sharing and reuse that can increase the collective benefits obtained from societal investment in science and engineering.

4.6. Limitation

The sample in this preliminary study included only a subset of the range of STEM disciplines, only one or two researchers from each of these disciplines, and only researchers from eastern U.S. research universities. Each interviewee reported observations and his/her own experiences from their personal research careers, so it is likely that the results are idiosyncratic for certain disciplines – and particularly those where there is substantial variation in sub-disciplinary practices. Therefore, the frequencies of each coding scheme would be limited in its interpretation. In future research, I need to include a more representative range of scholars and a more deliberate effort to obtain participants from a representative set of sub-disciplinary areas. Although the interview provides rich data, future research should also include mixed methods (e.g., surveys) in order to triangulate on the findings offered here. In addition, an objective snapshot of available repositories and metadata standards for presentation to informants could elicit more specific responses to why a researcher uses or does not use a particular data sharing resource. In addition, I focused in this study primarily on the motivations and challenges to *sharing data* rather than those associated with using deposited data. Although certain questions assessed both sides of the data sharing equation, I found that using other researchers' data is still new to many researchers.

4.7. Summary

This preliminary study shows three groups of data sharing factors including institutional influences, individual influences, and institutional resources. In terms of institutional factors, STEM researchers reported that pressures by funding agencies, journal publishers, private funding organizations, and their disciplinary influences affected their data sharing practice. In terms of individual motivation factors, researchers reported that perceived benefits (e.g. academic credits), efforts (e.g. annotation, organization), and risks (e.g. getting scooped) of data sharing influenced their data sharing. Lastly, in terms of institutional resources, researchers reported that internal capability (e.g. local IT support) and external capability (e.g. data repository) affected their data sharing. In addition, altruism emerged as an important factor influencing researchers' data sharing.

Results showed support for an institutional perspective on data sharing as well as an individual perspective for better understanding of scientists' data sharing behaviors. To have well-established data sharing practices, researchers need to have supportive institutional environments (e.g. data sharing structures, norms, policies), sufficient infrastructure (e.g. metadata standards, repositories), and positive attitudes toward data sharing (e.g., perceived benefits, efforts, risks). The results of this research synthesis were used to assist in the development of research model and the design of a survey that was distributed to diverse scientific disciplines at the main stage of this research.

Acknowledgement and Note:

The preliminary study in this chapter was guided by Dr. Jeffrey Stanton, and a partial writing in the discussion section was written by Dr. Stanton. The preliminary study has been published in the Journal of Computational Science Education (Volume 3, Issue 1).

5. Research Model and Hypotheses

A refined research model and its hypotheses are developed based on theories, previous literature, and the preliminary study. The conceptual model presented in the Chapter 3 provides an extensive map of scientists' data sharing behaviors according to the combination of institutional theory and the theory of planned behavior. However, this research focuses on selected research constructs by considering the results of preliminary study and prior studies, and this research develops its specific research model. The research model shows the complementary nature of the individual and institutional factors influencing scientists' data sharing behaviors.

5.1. Research Model

Based on the conceptual model, a refined research model is developed to explain and predict scientists' data sharing behaviors. This research model includes previous studies' findings and incorporates the findings from the preliminary study in this research. Drawing on theories, previous literature, and the preliminary study, 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.

This research focuses on selected research constructs based on the results of preliminary study and prior studies. The institutional factors include regulative pressures (from funding agencies and journals), normative pressures (from each discipline), and institutional resources (e.g. data repositories); individual factors including behavioral beliefs for attitude (i.e. perceived benefits, risks, and efforts toward data sharing behavior) and altruism. Since the research constructs of cultural-cognitive pressure in institutional theory and subjective norm in the theory of planned behavior were found minimally, those research constructs were removed for the final research model. Therefore, the final research model only considers regulative pressure, normative pressure, and institutional resource (i.e. resources-facilitating conditions as the external perceived behavioral control) at a discipline level, and it assesses behavioral beliefs for attitude and actual data sharing behavior at an individual level. By focusing on scientists' perceptions of benefits, risks, and efforts toward data sharing along with regulative and normative pressures, this study seeks to explore what combination of institutional and individual factors that influence scientists' decisions to share data with others. The Figure 5.1 below shows the research model for scientists' data sharing behaviors.

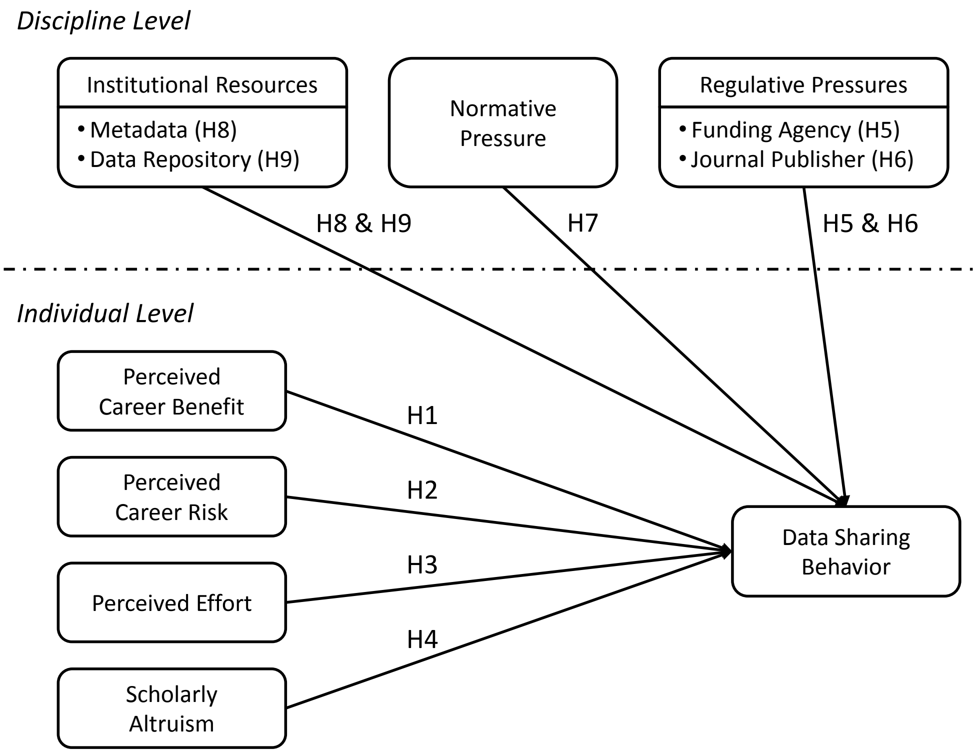


Figure 5.1 Research Model and Hypotheses (H) for Scientists' Data Sharing Behaviors

The multilevel model above shows how institutional and individual factors influence scientists' data sharing behaviors. For the institutional level factors, this research includes regulative pressures from funding agencies and journal publishers, normative pressure, and institutional resources (i.e. metadata and data repository); for the individual level factors, this research considers individual scientist's behavioral beliefs toward data sharing (i.e. perceived career benefit, perceived career risk, perceived effort) and scholarly altruism. This research eventually considers individual scientist's data sharing behavior as an outcome variable influenced by those institutional and individual factors. Scientists' data sharing behaviors can be best explained by considering both institutional and individual level factors together, and this research can shows how both institutional and individual factors cause scientists to make their decisions on data sharing. Each construct and related hypothesis is provided below.

5.2. Research Hypotheses

5.2.1. 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 and my preliminary study, I found that these three behavioral beliefs are the main individual level perceptions which either positively or negatively influence scientists' data sharing behaviors. Perceived career benefit would positively influence scientists' data sharing behaviors; however, both perceived career risk and perceived effort would negatively influence scientists' data sharing behaviors. Lastly, this research considers scholarly altruism as an important individual level factor influencing scientists' data sharing behaviors. Scholarly altruism would positively influences 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. Perceived career benefit is the value that scientists derive from demonstrating quality work, having more citations and credits, and eventually increasing their reputation and recognition of their work. Since scientists consider recognition and reputation to be valuable to their careers, they believe that sharing data can benefit their career by helping to increase their recognition and reputation.

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 et al. 2003); however, if they perceive low or no reward, they are unlikely to share their data with others (Sterling et al. 1990). In the context of knowledge sharing, perceived (career) benefits in the forms of recognition, reputation, and rewards were found to have significant influences on individuals' knowledge sharing attitudes and their intentions to share knowledge (Jones et al. 1997). My preliminary study also confirmed that scientists are willing to share their data because they perceive career benefits from data sharing (e.g. increased citation, possible credit, demonstration of quality work)**.** 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

The perceived career risk involved in data sharing would negatively influence scientists' data sharing behaviors. A risk refers to the natural probability of having an undesirable consequence. Prior studies defined perceived risk as the degree to which a person believes that his/her behavior has such as negative outcome (Conchar et al. 2004; Lee et al. 2009). In this study perceived career risk is defined as a scientist's belief about the potential uncertain negative outcomes from data sharing, which affect their career undesirably. 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. Based on my preliminary study, the potential negative outcomes of data sharing can be categorized into three groups including (1) losing control over data, (2) losing publication opportunities, and (3) getting scooped. These outcomes negatively influence scientists' academic careers.

Perceived risk has been studied in different areas, including online customers' perceived risk (Miyazaki et al. 2001; Shin 2008), consumer behavior (Pavlou 2003; Taylor 1974), organizations' technology adoption (Benlian et al. 2011), and information sharing (Awad et al. 2006; Posey et al. 2010). The concept of risk has sometimes been studied with regards to trust, which is a critical element of an organizational climate that facilitates knowledge utilization and exchange (Inkpen 1996; Roberts 2000). In the context of scientists' data sharing, prior studies identified diverse components of perceived (career) risk including losing publication opportunities (Reidpath et al. 2001; Savage et al. 2009; Stanley et al. 1988), protecting one's career (Campbell et al. 2002; Louis et al. 2002), and misuse of data (Borgman 2007; Cragin et al. 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

The perceived effort required to share data would negatively influence scientists' data sharing behaviors. Perceived effort refers to the degree to which a scientist believes that sharing data would require work (energy) and time. In regards to technology adoption studies, perceived effort corresponds to effort expectancy, "the degree of ease associated with the use of the technology" (Venkatesh et al. 2003), and perceived ease of use, "the degree to which a person believes that using a particular system would be free from effort" (Davis 1989). 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 et al. 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 et al. 2008). Some previous studies in knowledge sharing defined the concept of altruism as a form of intrinsic motivation (Cho et al. 2010; Hung et al. 2011a; Hung et al. 2011b; Lee et al. 2010), since altruism provides few tangible rewards, but offers psychological benefits such as satisfaction and enjoyment of helping others (Krebs 1975; Smith 1981). However, this research defines the concept of scholarly altruism by focusing on an individual's willingness to help others and contribute to the welfare of his or her community without expecting returns (Baytiyeh et al. 2010; Fehr et al. 2003; Fehr et al. 2006; Kankanhalli et al. 2005). The idea of intrinsic motivation was excluded in the concept of scholarly altruism. The preliminary study showed that in the context of scientists' data sharing, scholarly altruism motivates scientists to help other scientists save time and effort, allow others to find something missing from the original research, and help them contribute to scientific development in their research fields.

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 et al. 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 et al. 2005; Kim 2007). In the context of knowledge sharing, altruism was found to be an important factor influencing individuals to share their knowledge with others (Constant et al. 1996; Davenport et al. 1998; He et al. 2009; Hung et al. 2011a; Kankanhalli et al. 2005; Lin 2008). Those studies have showed that altruism has a significant influence on individuals' knowledge sharing attitudes and their intention to share knowledge (Cho et al. 2010; Constant et al. 1994; Lin 2007). My preliminary study also shows that scientists share their data based on their scholarly altruism. 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 availabilities of metadata standards and data repositories).

This research model does not consider the behavioral intention included in Ajzen's (1991) original model. The behavioral intention is assumed to capture individual motivational factors such as attitude, subjective norm, and perceived behavioral control (Ajzen 1991), and the intention to perform or not perform a behavior is an immediate determinant of the actual behavior (Ajzen et al. 1985). The construct of behavioral intention has been criticized because of its low ability to predict actual behavior (Burton-Jones et al. 2006; Jasperson et al. 2005; Kim et al. 2005; Limayem et al. 2007). Ajzen (1991) reported that the three predictors (i.e. attitude, subjective norm, perceived behavioral control) of intention can explain 50 percent of the variance in intention on average; however, intention only explains 26 percent of the variance in behavior on average (Ajzen 1991). In this research, the actual data sharing behavior was measured in order to examine the direct relationships between individual and institutional predictors and scientists' data sharing behaviors.

5.2.2. Institutional Level

Regulative Pressures (by Funding Agencies and Journal Publishers)

Governmental funding agencies and journal publishers exert regulative pressures on scientists regarding their data sharing behaviors. They 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 coercion from scientific funding agencies such as NSF and NIH, which are resource dominant organizations, so they need to comply to secure their own survival (Pfeffer et al. 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 et al. 2010; Teo et al. 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 et al. 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 et al. 2001; Zsidisin et al. 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.

Institutional Resources (Metadata Standard and Data Repository)

Institutional resources including metadata standards and data repositories in a discipline positively influence scientists' data sharing behaviors. The institutional resources which are already known as resource-facilitating conditions in prior studies would be important institutional level factors influencing scientists' data sharing. Resource facilitating conditions were originally studied as external behavioral controls in the theory of planned behavior (Ajzen 1991). Compared to self-efficacy (i.e. internal perceived behavioral control), which focuses on individual's own capability to perform a behavior, resource-facilitating conditions (i.e. external perceived behavioral control) is defined as individual judgments about the availability of facilitating resources and environments to perform a behavior (Ajzen 1991; Hsu et al. 2004; Taylor et al. 1995). In the context of scientists' data sharing, resource-facilitating conditions mean the availability of necessary resources including metadata standards and data repositories in a discipline for scientists' data sharing.

According to the theory of planned behavior, resources-facilitating conditions as the external perceived behavioral control influence an individual's attitude, intention, and his/her actual behavior (Ajzen 1991; Hsu et al. 2004; Taylor et al. 1995). In addition, 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 et al. 2003; So et al.

2005). Therefore, scientists' resource-facilitating conditions including metadata standards and data repositories would enhance scientists' data sharing behaviors.

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

H9: 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. In the next chapter, I present the research methodology and relevant issues for survey research.

5.3. Methodological Consideration

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). Therefore, this research employs a multilevel analysis in order to validate the research model and hypotheses developed above.

5.4. Summary

This research model explains and predicts scientists' data sharing behaviors. It includes previous studies' findings in data sharing and incorporates findings from my preliminary study. This research model identifies two groups of factors – individual influences and institutional influences, respectively – that influence scientists' data sharing behaviors. This research model shows the complementary nature of the individual and institutional factors influencing scientists' data sharing behaviors. 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.

Based on the research model developed from institutional theory and theory of planned behavior, this research proposes several hypotheses to be tested empirically. Those hypotheses focus on individual level and discipline level: At the individual level, this research examines whether perceived career benefit, perceived career risk, perceived effort, and scholarly altruism influence individual scientist’s data sharing behavior. At discipline level, this research examines whether regulative pressures by funding agencies and journal publishers and normative pressure in each discipline influence scientist’s data sharing behavior. Lastly, this research also examines whether institutional resources including metadata standards and data repositories influence scientist’s data sharing behavior.

6. Methodology

This chapter describes the procedure of survey method employed in this research as a main research method. The following sections contain the details of survey method, including research design; population and sampling; instrument development; and relevant reliability and validity issues. This research has created its own survey instrument through a series of steps. The instrument development section presents a procedure that includes item creation, scale development, and instrument testing. At the end of this chapter, I provide a data collection procedure and data analysis plan for the field survey conducted in diverse science and engineering disciplines.

6.1. Population and Sampling

6.1.1. Target Population

The target population of this research includes 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 can be identified from the scholar list in the Community of Science's (CoS) Scholar Database (http://pivot.cos.com), which provides a researcher profile directory in the world mainly from universities and colleges. The CoS scholar database provides the means to directly access the population of this research. Based on the list of scholars who are registered in U.S. academic institutions, scientists are randomly selected from STEM disciplines categorized in the CoS database.

The CoS database currently has the profile directory of over 3 million (3,188,174 as of 9/16/2012) scholars around the world in 15 major academic disciplines. The 15 major discipline categories include agriculture, allied health, applied science, architecture, arts, business, education, engineering, environmental science, humanities, law, mass communication, medicine, natural science, and social science. Scientists' profiles are created based on publicly available information, mainly from university websites and also user input. The original purpose of the CoS database is to help researchers find any potential collaborator across multiple disciplines based on topics of interest. The CoS database provides each scholar's profile information, including affiliation, expertise, publication and grant summary, communities, keywords, degrees, personal website, and contact information (address and email).

In the United States, there are 1,663,156 registered scholars in 15 major disciplines categorized by the CoS scholar database (as of 9/16/2012). By using query, I identified a total of 533,674 scholars in STEM disciplines (categorized by NSF discipline codes), including Engineering (67,146), Physical Sciences (52,996), Earth, Atmospheric, and Ocean Sciences (17,778), Computer Science (30,680), Agricultural Sciences (16,568), Biological Sciences (113,120), Psychology (25,677), Social Sciences (52,107), and Health Sciences (157,602). Each population of nine main STEM disciplines and 56 sub-disciplines can be found in the Appendix 8.2. The list of scholars in each discipline includes faculty members, post-doctoral researchers, and sometimes graduate student researchers. The sampling frame used in this research is close to the desired research target population. Based on the sampling frame, I can select the sample in each discipline by using random sampling method.

6.1.2. Sampling Plan

Sample Size

This research employs a multilevel analysis for its statistical analysis technique. In multilevel analysis, the sample size depends on the number of participants in one group and the number of groups. There is no concrete agreement about adequate sample size (i.e. number of groups and number of members in each group) for multilevel analyses (Raudenbush et al. 2002). Prior studies recommended a minimum of 30 to 50 groups with 20 to 30 members in each group as necessary for multilevel analysis (Bickel 2007; Heck et al. 1999; Hox 2002; Maas et al. 2005). In terms of the Level-1 sample size, scholars have suggested that a minimum of 20 observations in each group is required to have stable measurements for aggregated group-level variables (Hox 2002; Scherbaum et al. 2009). For the Level-2 sample size, scholars have recommended at least 10 groups necessary for each group-level predictor (Goldstein 2011; Raudenbush et al. 2002). In addition, scholars have argued that it is more important to increase the number of groups included for multilevel analysis, as opposed to the number of members in each group (Zhang et al. 2009).

This research planned to collect a sample size of at least 50 disciplines, with a minimum of 20 scientists per discipline according to the sample size recommendations of prior studies (Goldstein 2011; Hox 2002; Raudenbush et al. 2002; Scherbaum et al. 2009). Since this research has five Level-2 predictors (i.e. regulative pressures by funding agencies and journals, normative pressure, availabilities of metadata and data repositories), it is necessary to have at least 50 disciplines to detect Level-2 effects (Goldstein 2011; Raudenbush et al. 2002). Also, 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 et al. 2009). Therefore, the sample size required for this research should be equal to or greater than 1,000 scientists who belong to at least 50 disciplines, with a minimum of 20 scientists comprising each discipline. This sample size can allow conducting a valid multilevel statistical analysis.

Sampling Strategy

The sampling frame which I use for this research represents the target population, so the results of the sample can be generalized to the population. The survey participants were sampled based on a probability random sampling method. From the CoS scholar database, the potential participants were randomly selected from a panel of individual scientists who work in U.S. academic institutions, have occupational titles of faculty, researcher, and post-docs, and have Ph.D. degrees. Especially, potential participants are expected to have at least one publication based on research data generated in the last two years.

A pilot survey was conducted to understand the reliability and feasibility of the CoS scholar database for the sampling frame of this research. From the pilot survey distribution with 400 randomly selected potential participants, it was found that about 20.50 % (82 people) of the randomly-selected scientists in ecology were not usable because they do not have email addresses (11, 2.75%), or the email addresses provided are not valid (71, 17.75%). A total of 318 people (79.50%) were identified as potential survey participants and were asked to take the pilot survey. Among 318 potential participants, 34 people (10.69%) participated in the online survey (without any reminders), and 26 people out of 34 actual participants were found to be ecologists. In addition, it was found that some graduate students and staff members have incorrect titles, and are inappropriately registered as scientists.

Based on the pilot survey result above, the field survey needs to be distributed to about 300 potential participants in at least 50 disciplines to effectively secure a minimum of 20 valid scientists in each discipline. The pilot survey also shows that about one-fifth of the registered scientists in the CoS database were not reachable due to invalid email addresses. Therefore, for the final survey distribution, 400 people in each discipline should be randomly selected from the CoS scholar database in order to expect to have 300 potential participants in each discipline with valid email addresses.

Since there are a total of 533,674 registered scholars in nine main STEM disciplines and 56 sub-disciplines in the CoS scholar database (the disciplines of mathematics and statistics were excluded since their research focuses on theoretical works and usually does not generate any data), 400 people were randomly selected from 56 STEM sub-disciplines (except psychology). Since psychology has three sub-disciplines (clinical, non-clinical, and combined) according to NSF discipline codes, 1,200 people were randomly selected from the psychology discipline as categorized in the CoS scholar database. This resulted in 23,200 people randomly selected from 56 STEM disciplines. The detailed process of survey distribution was provided in the Section 5.5 Data Collection Procedure of this chapter.

6.2. Instrument Development

In this section, the process of survey instrument development is described. The development and validation of the survey instrument follows the prescribed set of steps proposed by Moore and Benbasat (1991). They laid out three stages of instrument development, including item creation, scale development, and instrument testing. The scale development procedure is shown in Figure 6.1. Scale development is necessary in this research because prior studies did not test their measurement items in scientists' data sharing context. In addition, this research developed new measurement items for some of the constructs through the scale development procedure.

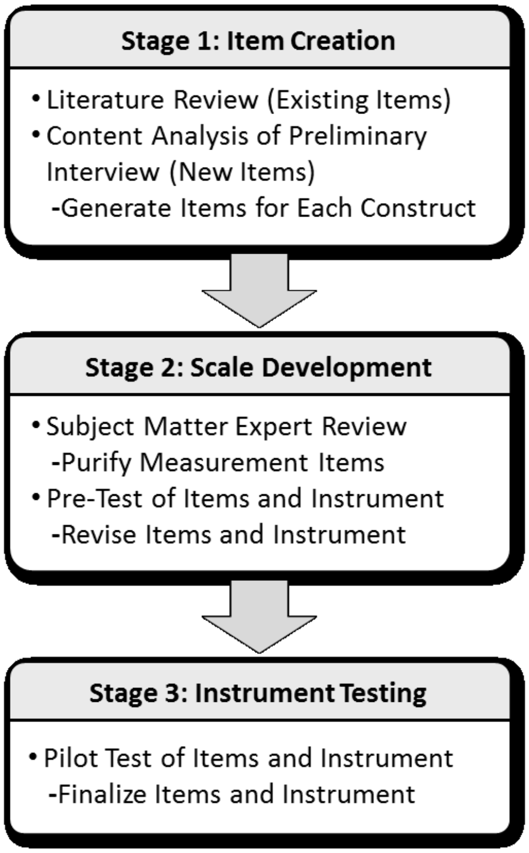

Stage 1: Item Creation
• Literature Review (Existing Items)
• Content Analysis of Preliminary Interview (New Items)
-Generate Items for Each Construct
Stage 2: Scale Development
• Subject Matter Expert Review
-Purify Measurement Items
• Pre-Test of Items and Instrument
-Revise Items and Instrument
Stage 3: Instrument Testing
• Pilot Test of Items and Instrument
-Finalize Items and Instrument

Figure 6.1 Scale Development Procedure

6.2.1. Stage 1: Item Creation

At the item creation stage, the initial measurement items were created based on prior literature and preliminary interviews. As the first step of item creation, each construct was defined according to the theoretical framework. Then, an extensive literature review was conducted to identify and evaluate the existing measurement items for each construct. In addition, new measurement items were generated based on the content analysis of preliminary interviews in order to fill out the gaps between existing measurement items and the constructs studied in this research. The definition of each construct was provided in the Table 6.1 below, and the literature review was presented in Chapter 2.

Construct	Definition	Source
Perceived Career Benefit	The degree to which a scientist believes that sharing data could provide rewards such as recognition and reputation through acknowledgements, citations, and sometimes authorships	(Bock et al. 2005)
Perceived Career Risk	A scientist's belief about the potential uncertain negative outcomes from data sharing, which affect their career undesirably	(Featherman et al. 2003)
Perceived Effort	The degree to which a scientist believes that sharing data would require work (energy) and time	(Davis et al. 1989)
Scholarly Altruism	The degree to which a person is willing to work to increase others' welfare without expecting any returns	(Hsu et al. 2008)
Regulative Pressure by Funding Agency	Coercive aspects of funding agencies which regulate and constrain scientists' data sharing behaviors	(Scott 2001)
Regulative Pressure by Journals	Coercive aspects of journals, which regulate and constrain scientists' data sharing behaviors	(Scott 2001)
Normative Pressure	The legitimizing means that stem from collective expectations in a scientific discipline	(Scott 2001)
Metadata	A set of data that provides information about one or more aspects of the original research data	(Venkatesh et al. 2003)

Data Repository	A digital archive where scientists can deposit their data of published articles and download other researchers' data	(Venkatesh et al. 2003)
Data Sharing Behavior	The extent to which scientists provide their research data and information related to their published articles with other scientists by depositing them into data repositories and providing them upon request	(Ajzen 1991)

Table 6.1 Definitions of Each Construct in the Research Model

The pertinent measurement items in prior literature were reviewed for coverage, reliability, and validity. Most of the measurement items were adapted for this research with minor modifications. In the selection of initial items, if similar items appeared in different sources, only well-tested items were adopted for the pre-test of the initial items (Moore et al. 1991). However, any slightly redundant items were included for subject matter experts to review and pretest in the scale development stage (DeVellis 2003). The complementary use of the measurement items from multiple sources would increase both breadth and validity of the instrument (DeVellis 2003). At this item creation stage, three to four times more items than the final survey items were developed, then those items were reviewed by the subject matter experts and pretested by a small sample of target population at the scale development stage.

Research Constructs		From Literature	Newly Created	Total Items
Discipline Level Predictors	Regulative Pressure by Funding Agencies	9	2	11
	Regulative Pressure by Journal Publishers	9	2	11
	Normative Pressure by Disciplines	10	2	12
	Metadata	9	2	11
	Data Repository	9	2	11
Individual Level Predictors	Perceived Career Benefit	10	3	13
	Perceived Career Risk	11	2	13
	Perceived Effort	9	4	13
	Scholarly Altruism	12	7	19
DV	Data Sharing Behavior	5	6	11
Total		93	32	125

Table 6.2 Numbers of Initial Items Adapted from Literature and Newly Created

A total of 125 initial measurement items for 10 constructs were identified from prior literature (93 items) and newly developed based on the content analysis of the preliminary interviews (32 items). While most of the measurement items were adapted from prior studies on institutional theory (Kostova et al. 2002; Son et al. 2007; Teo et al. 2003) and knowledge sharing (Baytiyeh et al. 2010; Bock et al. 2005; Kankanhalli et al. 2005; McLure Wasko et al. 2000), and technology adoption (Davis 1989; Davis et al. 1989; Taylor et al. 1995; Thompson et al. 1991; Venkatesh et al. 2003), new measurement items were developed in areas of limited numbers of measurement items. In particular, some of the scholarly altruism items were newly created for this study, based on the theoretical literature and the preliminary study (Batson 1991; Fehr et al. 2003; Fehr et al. 2006). The content analysis of the preliminary interviews not only compensated, but also validated the measurement items from the prior literature. The numbers of initial items adapted from literature and newly created for research constructs are shown in Table 6.2.

6.2.2. Stage 2: Scale Development

Subject Matter Expert Review

At the scale development stage, a panel of judges (who are the Subject Matter Experts (SMEs) from diverse scientific disciplines) reviewed and purified the initial measurement items generated at the item creation stage. The objectives of this scale development stage include (1) the evaluation of the construct validity of the items developed initially; and (2) refinement of the ambiguous items after the initial item creation (Moore et al. 1991). In addition, the survey instrument needs to be understood by scientists in diverse disciplines, so it was assured that the SMEs from different disciplines understood the survey questionnaires by producing more generalized statements.

The panel of judges was comprised of six faculty members and two post-doctoral researchers in the disciplines of biology, ecology (post-doc), chemistry (two professors), computer science (post-doc), environmental engineering, industrial engineering, and electrical engineering. They were provided with the definitions of constructs and asked to examine how well the initial items represented each construct. They evaluated the initial measurement items based on the definitions of the constructs, and provided feedback regarding the appropriateness of the items, sentence structure, and phrasing according to their research contexts. In particular, the review by the panel of SMEs was utilized to improve the clarity, readability, understandability, and appropriateness of the measurement items.

According to the feedback and comments of the panel of judges, I removed and modified some of the items which were redundant, did not cover the meaning of each construct, and mislead survey participants with differing interpretations. However, some of the redundant and similar items were included for the later pretest in order to check their reliability and validity with other items in each construct. Also, the feedback from the panel of judges resulted in modifying the number of scale points and the survey instruction and layout. After this purification and refinement process, the total number of initial items, 125, was substantially reduced, to 77. The number of items initially created and the number of items remaining for each construct are shown in Table 6.3. A full list of the purified and refined items can be found in the Appendix 8.3.

Research Constructs		Number of Initial Items	Number of Pretest Items
Discipline Level Predictors	Regulative Pressure by Funding Agencies	11	8
	Regulative Pressure by Journal Publishers	11	8
	Normative Pressure by Disciplines	12	8
	Metadata	11	7
	Data Repository	11	7
Individual Level Predictors	Perceived Career Benefit	13	10
	Perceived Career Risk	13	8
	Perceived Effort	13	8
	Scholarly Altruism	19	8
DV	Data Sharing Behavior	11	5
Total		125	77

Table 6.3 Numbers of items for each construct before and after SME review

Pre-Test of Items and Instrument

A pretest of the purified items from the SME review was conducted to revise and refine the measurement items by using reliability analysis and feedback from individual scientists representing the target population. A pretest is desirable in a survey study since the survey participants only can answer the survey questions and items provided in a survey questionnaire (Dillman 2007). The pretest also helped to reduce the number of survey items to be included in the field survey. Any items which had measurement errors or did not share the core value with other items in each construct were removed at this stage.

Main-Discipline	Sub-Discipline	Number of Respondents	Percentage of Respondents
Engineering	Aerospace Engineering	1	3.45%
	Biomedical Engineering	3	10.34%
	Civil Engineering	1	3.45%
	Electrical Engineering	1	3.45%
Physical Sciences	Chemistry	1	3.45%
	Physics	1	3.45%
Earth Sciences	Geosciences	3	10.34%
Mathematical Sciences	Mathematics	1	3.45%
Computer Science	Computer Science	2	6.90%
Agricultural Sciences	Forestry	1	3.45%
Biological Sciences	Biology	1	3.45%
	Cell and Molecular Biology	2	6.90%
	Ecology	3	10.34%
	Genetics	1	3.45%
	Pathology	1	3.45%
Psychology	Clinical Psychology	2	6.90%
	Psychology, Except Clinical	4	13.79%
Total		29	100.00%

Table 6.4 Research Disciplines of Pre-Test Participants

The pretest was performed with scientists including faculty members and post-doctoral researchers in STEM disciplines at a research institution in the eastern U.S. The pretest instrument was sent to 268 potential participants by email on October 23, 2012. The email message for this pretest included information about the purposes of this research, the pretest survey, and the online survey link. Only one reminder was sent after one week (on October 30, 2012). A total of 29 scientists participated in the pretest survey either partially or fully. The response rate of this pretest was low (10.82%) because of the potential that participants might learn that the survey instrument was not an actual survey and so might decide not to participate in the pretest (Dillman 2007). The discipline information of pretest participants is shown in Table 6.4, and Table 6.5 below shows the demographics of pretest participants.

Profile Category		Number of Respondents	Percentage of Respondents
Gender	Female	11	37.93%
	Male	18	62.07%
Age	25-34	4	13.79%
	35-44	13	44.83%
	45-54	0	0.00%
	55-64	7	24.14%
	65+	5	17.24%
Education	PhD/Doctoral Degree	29	100%
Position	Assistant Professor	10	34.48%
	Associate Professor	8	27.59%
	Full Professor	7	24.14%
	Professor Emeritus	1	3.45%
	Lecturer/Instructor	1	3.45%
	Post-Doctoral Fellow	1	3.45%
	Researcher	1	3.45%
Total		29	100%

Table 6.5 Demographics of Pretest Participants

At the pretest stage, 10 constructs containing 77 items were pretested for reliability of measurement. The reliability of the refined items from the SME review was assessed using the item-to-total correlation coefficient and Cronbach’s alpha. Item-to-total correlation refers to the relationship of the selected item with the sum of the other items. The items whose item-to-total correlation is less than .6 were dropped or reworded (Nunnally et al. 1994) since those items provide low explanation power and attenuate the overall reliability of the items for each construct (Nunnally et al. 1994). Also, any items whose Cronbach’s alpha if the item was deleted is larger than overall Cronbach’s alpha was removed. Cronbach’s alpha is commonly used to measure reliability. The value of Cronbach’s alpha greater than .7 can be considered as a good measure (Nunnally et al. 1994). In this stage, some similar items from different prior studies were carefully examined, so any redundant and similar items were removed or reworded. Through this pretest process, only three to four items were selected for each construct in order to minimize the response time in the final instrument. Cronbach’s alpha for original items and Cronbach’s alpha for selected items for each construct are shown in Table 6.6.

Variable	Number of Original Items	Cronbach's α for Original Items	Number of Selected Items	Cronbach's α for Selected Items
Regulative Pressure by Funding Agency	8	.874	4	.809
Regulative Pressure by Journals	8	.908	4	.885
Normative Pressure by Disciplines	8	.926	4	.866
Metadata	7	.842	3	.820
Data Repository	7	.809	3	.851
Perceived Career Benefit	10	.913	4	.859
Perceived Career Risk	8	.896	4	.843
Perceived Effort	8	.905	4	.887
Scholarly Altruism	8	.856	6	.831
Data Sharing Behavior	5	N/A	5	N/A
Total	77		41	

Table 6.6 Reliability of Each Independent Variable (Pretest: n=29)

After the pretest for reliability, 36 items were removed, and only 41 items were retained for the pilot testing and final field distribution. The detailed procedure and explanation of refining items for each construct in the pretest stage was presented in the Appendix 8.4. The list of items which were deleted in this stage can be found in the same Appendix.

Revisions of Item and Instrument

From the pretest, any potential problems were identified in the survey instrument. The respondents provided their feedback regarding the instruction, format of the survey, measurement scale, and wording of the items. There were some significant changes made at this stage. They included: (1) Any redundant measurement items were removed, leaving only key measurement items in each scale; (2) Questions were grouped into five parts including introduction, institutional pressure, individual perceptions, their data sharing behaviors, and the demographic information; (3) Several questions were included to identify scholars who generate actual research data, such as “Do you produce actual research data?," aimed at identifying the scientists who generate scientific research data; (4) The items measuring data sharing behaviors were updated logically to address diverse types of data sharing behaviors. Also, (5) a seven-point Likert scale was selected for all measurement items for consistency purposes. The measurement scales range 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. Lastly (6), the overall clarity and comprehensibility were improved by feedback from pretest survey participants (See the Appendix 8.5 for details).

6.2.3. Stage 3: Instrument Testing

Pilot-Test

At the instrument testing stage, a pilot test of the survey instrument from the prior scale development stage was conducted with a representative sample out of the target population. The main objective of this pilot test was to ensure that "the various scales demonstrate the appropriate levels of reliability" (Moore et al. 1991). Since the survey instrument in this research uses multiple measurement items, reliability of the measurement items for each construct is critical. For reliability assessment, this research employs Cronbach's alpha and item-to-total correlations.

Out of 4,006 scientists listed in the discipline of ecology in the CoS scholar database, 400 scientists were randomly selected for this pilot test. The pilot test instrument was distributed by email on November 12, 2012, and no reminder was sent. The email message included introduction to and purpose of the survey, and the link to the pilot survey. Another purpose of the pilot test was to assess whether contact information listed in the CoS scholar directory is reliable, and how many scientists chosen from the CoS directory would respond to this survey. There were 82 people (20.50%) who could not be reached because they lacked email addresses (11, 2.75%); or the emails listed were returned due to invalid addresses (71, 17.75%). A few people responded regarding their ineligibility to be considered in this pilot survey, either because they are retired and do not produce any research data, or because they are not scientists. Therefore, 82 out of 400 were removed from the pilot sample, and 318 out of 400 email messages were delivered to the potential participants. A total of 36 submissions were recorded on the survey website, and out of the 36 submissions, there were 34 valid responses used for the data analysis of the pilot test. The profiles of the pilot test sample are shown in Table 6.7.

Profile Category		Number of Respondents	Percentage of Respondents
Discipline	Ecology	26	76.47%
	Forestry	3	8.82%
	Plant Sciences	2	5.88%
	Biophysics	1	2.94%
	Microbiology	1	2.94%
	Biology	1	2.94%
Gender	Male	21	61.76%
	Female	11	32.35%
	Missing	2	5.88%
Age	25-34	7	20.59%
	35-44	5	14.71%
	45-54	9	26.47%
	55-64	11	32.35%
	65+	2	5.88%
Education	Bachelor’s Degree	3	8.82%
	Master’s Degree	2	5.88%
	PhD/Doctoral Degree	29	85.29%
Position	Graduate Student	5	14.71%
	Post-Doctoral Fellow	1	2.94%
	Researcher	6	17.65%
	Assistant Professor	5	14.71%
	Associate Professor	5	14.71%
	Full Professor	8	23.53%

Professor Emeritus	3	8.82%
Other	1	2.94%
Total	34	100%

Table 6.7 Demographics of Pilot-Test Participants

Pilot-Test Analysis

The psychometric properties of the scales in this pilot instrument were evaluated by using reliability measures. For the reliability measure, the pilot study employed Cronbach's alpha and item-to-total correlations. Cronbach's alpha ranged between .806 (Regulative Pressure by Funding Agencies) and .970 (Scholarly Altruism). The item-to-total correlations of the measurement items ranged between .525 to .956, which are above .50 (Doll et al. 1988; Netemeyer et al. 1996). The reliability values including Cronbach's alpha and item-to-total correlation based on the pilot test are shown in Table 6.8.

Variable	Number of Items	Cronbach's alpha	Item-to-Total Correlation	Number of Cases Used
Regulative Pressure by Funding Agencies	4	.806	.525 - .794	30
Regulative Pressure by Journals	4	.946	.805 - .918	30
Normative Pressure by Disciplines	4	.834	.546 - .772	33
Metadata	3	.928	.756 - .907	30
Data Repository	3	.933	.838 - .885	32
Perceived Career Benefit	4	.892	.560 - .863	33
Perceived Career Risk	4	.894	.772 - .826	34
Perceived Effort	4	.906	.726 - .808	34
Scholarly Altruism	6	.970	.817 - .956	31

Table 6.8 Reliability Values for Pilot Test (n=34)

Recommended Changes to the Survey Instrument

The reliability values on Table 6.8 above show that all the constructs are satisfactory in terms of their construct reliability. A few minor changes were made after this pilot testing stage because any significant changes would influence the reliability and validity of the items for the final study. The online survey system recorded the time spent to complete this survey, and it was found that 7-10 minutes was taken to complete the pilot survey. Although this study planned to avoid student scientists, retired scientists, and any scientists who work outside academic institutions; it was found that the CoS scholar database included student and retired scientists and non-academic scientists registered as scientists with incorrect titles. Therefore, additional demographic survey questions about scientists' job titles, educational background, and work sector were included in the final survey in order to identify valid participants for this research. This survey instrument was distributed to the rest of the scientists in diverse scientific disciplines in the final survey distribution.

6.2.4. Measurement of Constructs

The theoretical framework was translated into measurements of constructs. The measurement scales were refined and validated through the prior instrument development procedure. 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. In 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, which is more reliable than using a single or two-item measurement (Fabrigar et al. 1999; Rakov et al. 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 which best describes their level of agreement in the statements.

Since this research employs a multilevel model, institutional level constructs need 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 et al. 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 can examine the extent to which those shared property constructs are shared by individual scientists in a same discipline (Klein et al. 2000). The measurement items for each construct and its sources are indicated in Table 6.9.

Construct	Items	Sources
Regulative Pressure by Funding Agencies	• Data sharing is mandated by the policy of public funding agencies. • Data sharing policy of public funding agencies is enforced. • Public funding agencies require researchers to share data. • Public funding agencies can penalize researchers if they do not share data.	(Kostova et al. 2002) (Teo et al. 2003)
Regulative Pressure by Journals	• Data sharing is mandated by journals' policy. • Data sharing policy of journals is enforced. • Journals require researchers to share data. • Journals can penalize researchers if they do not share data.	(Kostova et al. 2002) (Teo et al. 2003)
Normative Pressure	• It is expected that researchers would share data. • Researchers care a great deal about data sharing. • Researchers share data even if not required by policies. • Many researchers are currently participating in data sharing.	(Kostova et al. 2002) (Son et al. 2007)
Metadata	• Researchers can easily access metadata. • Metadata are available for researchers to share data. • Researchers have the metadata necessary to share data.	(Thompson et al. 1991) (Taylor et al. 1995) (Venkatesh et al. 2003)
Data Repository	• Researchers can easily access data repositories. • Data repositories are available for researchers to share data. • Researchers have the data repositories necessary to share data.	
Perceived Career Benefit	• I can earn academic credit such as more citations by sharing data. • Data sharing would enhance my academic recognition. • Data sharing would improve my status in a research community. • Data sharing would be helpful in my academic career.	(McLure Wasko et al. 2000) (Bock et al. 2005)

Perceived Career Risk	• There is a high probability of losing publication opportunities if I share data. • Data sharing may cause my research ideas to be stolen by other researchers. • My shared data may be misused or misinterpreted by other researchers. • I believe that the overall riskiness of data sharing is high.	(Featherman et al. 2003) (Pavlou 2003)
Perceived Effort	• Sharing data involves too much time for me (e.g. to organize/annotate). • I need to make a significant effort to share data. • I would find data sharing difficult to do. • Overall, data sharing requires a significant amount of time and effort.	(Davis 1989) (Davis et al. 1989) (Thompson et al. 1991)
Scholarly Altruism	• I am willing to help other researchers by sharing data. • I would share data so that other researchers can conduct their research more easily. • I would share data so that other researchers can utilize it for their research. • I would share data to support open scientific research. • I would share data to contribute to better scientific research. • I would share data to help improve the quality of scientific research.	(Kankanhalli et al. 2005) (Baytiyeh et al. 2010) Newly Developed
Data Sharing Behavior	• How frequently have you deposited your data into disciplinary data repositories for every article? • How frequently have you deposited your data into institutional data repositories for every article? • How frequently have you uploaded your data into public Web spaces for every article? • How frequently have you provided access to your data by publishing supplement materials for every article? • How frequently have you responded to the data sharing request(s) by providing data via personal communication methods (e.g. email)?	Newly Developed

Table 6.9 Measurement Items for Research Constructs

6.3. Reliability and Validity

Reliability

This research considers the issues of reliability and validity (including content and construct validities). Since this research employs a survey as a main research method, issues of measurement reliability and validity are all important. Reliability includes both test-retest reliability and internal consistency of the items. Reliability is a precondition for securing measurement validity (Schutt 2006). Test-retest reliability refers to the extent to which a measure procedure yields consistent outcomes at different timeframes (Schutt 2006), and the internal consistency (which is also called inter-item reliability), refers to the extent to which multiple measures are consistent towards the same concept. This research ensures reliability in terms of test-retest issue and internal consistency by using well-developed items and performing instrument development procedures (i.e. item creation, scale development, instrument testing). Also, reliability assessment for each construct was conducted by checking internal consistency of variables. In terms of statistical methods, this research uses the Cronbach's alpha as the internal consistency (inter-item reliability) measure indicator (Schutt 2006).

Content Validity

Content validity refers to the extent to which the items cover the full range of the concept (Schutt 2006). Content validity can be ensured by reasonable instrument construction and representative items (Ragin 1994). Content validity was ensured by adapting the majority of the survey items from previous studies through in-depth literature review. In addition, content validity was warranted by presenting the survey items to a panel of judges who are the Subject Matter Experts (SMEs) from diverse scientific disciplines. At the scale development stage, eight SMEs were provided with the refined version of measurement items of 10 constructs, and they were asked to examine the survey items in terms of appropriateness and completeness of the measurements for each construct (Schutt 2006). Some of the items were modified to accommodate the recommendations and suggestions by the SMEs.

Construct Validity

Construct validity refers to the extent to which a set of items in a survey correctly operationalize the concept needing to be studied based on a theory (Schutt 2006). There are two approaches to construct validation: convergent validity and discriminant validity. Convergent validity refers to the extent to which one measure of a concept is similar to other measures of the same concept (Schutt 2006). Discriminant validity refers to the extent to which a measure of a concept is different from other measures of other concepts (Schutt 2006). In order to ensure construct validity, this research employs multiple items to measure each construct, and the survey items are adapted from the supportive literature. Construct validity was also warranted by the eight SMEs who reviewed the survey items in terms of convergent and discriminant validities. In terms of statistical method, the construct validity is evaluated by conducting factor analysis, to show whether common factors appear in multiple underlying items.

6.4. Data Collection Procedure

This section presents the data collection procedure of the survey study. Since this research involves human subjects, Institutional Review Board (IRB) approval was granted prior to data collection (i.e. interviews, survey). This research was approved by the IRB at Syracuse University, and the IRB documents are attached in the Appendix 8.14. The IRB allowed me to conduct preliminary interviews and pretest surveys with scientists (mostly faculty members) at Syracuse University, State University of New York – College of Environmental Science and Forestry, and Cornell University (additional approval was made from the IRB at Cornell University to recruit only the preliminary interview participants). The IRB at Syracuse University also allowed me 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 me to perform the random sampling using their scholar database for only the purpose of this research.

A PHP Web program was developed to randomly select the sample from the CoS scholar database. The Amazon Web Services was used to set up the PHP program to communicate with the CoS scholar database. By retrieving the scholar list from the database, each scholar's name, email address, discipline, affiliated institution and department, and position were recorded into the sample database in the Amazon Web server. Exactly 400 people were randomly selected from each of 56 STEM disciplines (except psychology – where 1,200 people were randomly selected from that discipline) based on the criteria of those professionals working in U.S. (academic) institutions and having Ph.D. degrees. The occupational title criterion was left open since some of the scientists' job titles were either missing or incorrect. The randomly retrieved scholar list was saved in the Web database at Amazon Web Services first, then it was downloaded as a DBF file to be used for field survey distribution.

A total of 23,200 people in 56 STEM disciplines were retrieved from the random sampling procedure above. By examining the retrieved scholar list, 1,369 people (5.90%) were found not to have any email addresses, and 42 people (0.18%) were removed because their email addresses were redundant. The initial email message introducing this research, the researcher, and the eligibility of this study was sent to those remaining 21,789 people on November 15, 2012. The initial email message is included in the Appendix 8.6. After the initial email was distributed, 5,036 email messages (21.71%) were returned and not delivered due to incorrect and invalid email addresses. Therefore, 1,411 ineligible people (due to no email address or redundant email addresses) and 5,036 invalid email addresses were removed from the distribution of the field survey instrument, and 16,753 out of 23,200 people from the random sampling were identified as potential survey participants. The result of random sampling and initial message distribution is summarized in Table 6.10.

Category		Frequency	Percentage
Number of Random Sample		23,200	100.00%
Excluded Sample	Email Missing	1,369	5.90%
	Redundant Email	42	0.18%
	Returned Email	5,036	21.71%
Number of Adjusted Sample		16,753	72.21%

Table 6.10 Result of Random Sampling and Initial Message Distribution

This led to 16,753 potential survey participants of 56 disciplines who would receive the following messages with the survey link. The survey questionnaire was created and distributed to individual scientists by using SurveyGizmo (http://www.surveygizmo.com). The student version of SurveyGizmo service allows setting up online surveys and collecting unlimited responses. The online survey questionnaire consists of research introduction and purpose, specific questions to measure the constructs, and respondents' demographic information. This online survey presents 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 were 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.

6.5. Data Analysis Plan

This research employs a combination of several statistical analyses techniques for the survey data collected. Surveys usually produce a quantified description for a defined variable, and many times they can show the relationship between variables based on statistical data analysis (Schutt 2006). Statistical data analysis methods help to make survey results generalizable into a large population, but statistical analysis methods do not provide such detailed explanations (Punch 2005). The survey data in this research have multiple measures for each construct, and also include some demographic and discipline information of the participants. The survey data are analyzed for descriptive statistics, reliability and validity analysis, and multilevel analysis.

This research uses Cronbach's alpha and factor analysis for the scale assessment. Since survey is the main research method, it is important to assess the reliability and validity of the measurement items. Cronbach's alpha is used to evaluate the reliability of the items for each construct. The construct validity was evaluated by using principal component factor analysis, which assesses the extent to which indicators specified for each measure refer to the same conceptual construct. Both convergent validity and discriminant validity can be assessed using principal component factor analysis approach. In addition, the one-way Analysis of Variance (ANOVA) test was conducted to examine nonresponse bias by comparing early and late respondents.

A multilevel regression analysis (also called Hierarchical Linear Modeling) is utilized as a main data analysis method to test the research hypotheses and answer the research questions. The hierarchical data collected from survey allows a multilevel analysis with scientists nested within their disciplines. The multilevel analysis allows investigating the nested nature of "scientists with disciplines" by simultaneously examining both discipline- and individual-level influences on scientists' data sharing behaviors. Before the multilevel analysis, the Intraclass Correlation Coefficients (ICCs) and r_{wg} statistics are used to assess whether the disciplinary-level variables are properly aggregated to the group level of analysis.

6.6. Summary

This chapter covers the procedure of survey method employed in this research as a main research method. The theoretical framework is translated into the measurements of constructs. The survey method can help to examine the constructs and hypothesized relationships of the scientists' data sharing model. By conducting the survey in diverse science and engineering disciplines, this research can validate the scientists' data sharing model and answer the research questions. The survey method can produce more generalized results about scientists' data sharing behaviors across different disciplines.

This research has created its own survey instrument through a series of steps including item creation, scale development, and instrument testing. At the item creation stage, the initial measurement items were created based on the prior literature and the preliminary interviews. At the scale development stage, a panel of judges reviewed and purified the initial measurement items, and the refined items were pre-tested by potential survey participants. At the instrument testing stage, a pilot test of the survey instrument from the prior scale development stage was conducted with a representative sample out of the target population.

The target population of this research includes faculty members and post-doctoral researchers in the U.S. academic institutions who belong to STEM disciplines. The sampling frame of this research is identified from the scholar list in the CoS scholar database. A total of 16,753 people of 56 disciplines were randomly selected and identified as potential survey participants. In order to analyze the survey data collected, this research uses a variety of statistical techniques including Cronbach's alpha, principal component factor analysis, ANOVA, and multilevel analysis. Especially, a multilevel regression analysis is utilized as a main data analysis method to test the research hypotheses and answer the research questions.

7. Survey Data Analysis and Results

This chapter provides survey data analysis and results. The data collection procedure, including data cleaning and preparation, is presented at the beginning, followed by a report on the demographics of survey participants. The next section covers the scale assessments in terms of reliability and validity of the measurement items. Then, a summary of the research's data aggregation and an evaluation of the assumptions of multilevel analysis are presented. Lastly, the results of multilevel analysis are presented according to the three-step multilevel modeling procedure, and the findings for the hypothesized relationships are provided. The next chapter discusses the results presented in this chapter and provides the implications of this research. The data analysis procedure in this chapter is summarized in Figure 7.1:

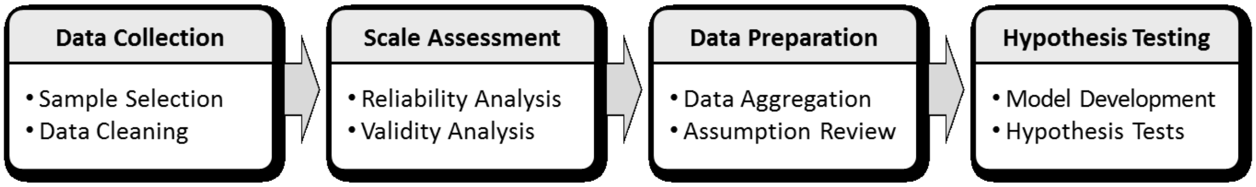


Figure 7.1 Data Analysis Procedure

7.1. Data Collection

7.1.1. Data Collection Results

The final field survey instrument was distributed to the 16,753 potential survey participants in 56 STEM disciplines by email on November 19, 2012. Those 16,753 potential participants were randomly selected from the CoS scholar database. They already received the initial email sent to introduce the research and survey, and to identify potential survey participants with valid email addresses. The email messages with the final field survey instrument were sent by using Outlook with mail-merge function. The email messages included an introduction, a description for the purpose of the survey, and a link to the online survey plus the online survey questionnaire. The questionnaire consisted of a brief research introduction and purpose statement, plus specific questions to measure the constructs as well as demographic questions.

Two reminders were sent, on December 17, 2012 and January 14, 2013, in order to encourage participation in the survey. These follow-up messages were needed to increase the response rate (Babbie 1990; Dillman 2007). After receiving 1,926 responses from the main survey by December 16, 2012, the first reminder was sent on December 17, 2012 to the same potential participants in the final field survey, except those who indicated they wanted to opt out and those who were not eligible for this survey (due to retirement, student scientist, non-scientist, returned email reasons). An additional 587 responses were received by January 13, 2013 after the first reminder (2,513 responses in total). The second and last reminder was sent on January 14, 2013 to the non-responding individuals. (Those who participated in the survey, and those who indicated they wanted to opt out were excluded from the last reminder.) After that reminder was sent, an additional 161 responses were received until the online survey was closed on February 15, 2013, with 2,674 responses in total.

Message Type	Date Sent	Number of Responses	Number of Accumulated Responses
First Email	11/19/2012	1,926	1,926
1st Reminder	12/17/2012	587	2,513
2nd Reminder	1/14/2013	161	2,674

Table 7.1 Number of Responses Received by Each Message Distribution

The numbers of responses received by each message distribution are shown in Table 7.1. The three email messages, including the first email message and the two reminders, are included in the Appendix 8.6, and the final survey instrument is included in the Appendix 8.7.

Although potential survey participants were refined through the initial email message distribution, there were still 197 returned emails (1.18%) because of incorrect email addresses. A total of 391 people responded that they were not eligible to participate in the survey due to retirement (252, 1.50%), student scientists (87, 0.52%), and non-scientists (52, 0.31%). Therefore, 588 out of 16,573 final survey recipients (3.51%) were removed from the response rate calculation, and a total of 16,165 participants (96.49%) received the email messages of the final field survey instrument. In addition, some scientists (464, 2.76%) replied that they did not want to participate in the following survey (I did not sent any further emails to this group, but I counted them as valid potential participants for the response rate calculation). The summary of the field survey distribution result is indicated in Table 7.2.

Number of Email with Survey Link Sent	16,753	100.00%
Returned Email (Not Delivered)	197	1.18%
Retired (by Reply)	252	1.50%
Student (by Reply)	87	0.52%
Not Scientist (by Reply)	52	0.31%
Adjusted Sample Size	16,165	96.49%
(Note) Opt Out	462	(2.76%)

Table 7.2 Summary of the Field Survey Distribution Results

The online survey on the SurveyGizmo website was accessible for the invited scientists for three months, from November 19, 2012 to February 15, 2013. A total of 1,926 responses were received after the first email was sent, and 587 additional responses were received after the second email (first reminder) was sent. The other 161 additional responses were received after the third email (second reminder) was sent. On February 15, 2013, the survey link was deactivated when there were no more responses during the week. 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 data analysis, and 204 responses were removed because those responses were missing more than 20% of answers and/or the answers regarding participants' data sharing behaviors, which is critical for the data analysis. A total of 2,470 responses remained as valid survey submissions. In this research, the sample size was adjusted from 16,753 (original sample size) to 16,165 (adjusted sample size) due to returned email (197), retirement (252), student (87), and non-scientists (52) (Pinelli 1991). This led to the response rate of 15.28% (2,470 valid responses out of 16,165 adjusted potential participants).

Main Disciplines	Survey Distributed	Valid Response Received	Response Rate
Engineering	2,831	356	12.58%
Physical Sciences	820	142	17.32%
Earth, Atmospheric, and Ocean Sciences	912	193	21.16%
Mathematical Sciences	-	17	-
Computer Science	305	25	8.20%
Agricultural Sciences	1,895	285	15.04%
Biological Sciences	4,338	789	18.19%
Psychology	838	95	11.34%
Social Sciences	1,447	266	18.38%
Health Fields	2,779	230	8.28%
Other Disciplines		49	
Missing		23	
Total	16,165	2,470	15.28%

Table 7.3 Response Rates by Disciplines

The response rates by main STEM discipline are shown in Table 7.3, and the response rates by specific STEM disciplines is included in the Appendix 8.8. The demographics and the disciplines of survey respondents (before the selection process) are included in the Appendix 8.9 and 8.10.

7.1.2. Inclusion and Exclusion Criteria

This research has a strict inclusion and exclusion criteria regarding its sample. This research originally planned to collect a panel of individual scientists who (1) work in the U.S. academic institutions, (2) have their Ph.D. degrees, and (3) hold occupational titles of faculty, researchers, and post-docs. In addition, this research only included (4) the scientists who are currently research-active and who produce their own data which can be shared with other scientists. The respondents who met the criteria above can answer the survey questions easily, and this increases the reliability and validity of the measurement items (Babbie 1990).

Another criterion used for the sample selection is the number of qualified respondents in each discipline who meet the above criteria. Since this research utilizes a multilevel analysis by aggregating individuals' responses in each discipline to group-level variables, it is necessary to have a minimum of 20 observations per discipline in order to ensure reliable measures for the group-level variables (Hox 2002; Scherbaum et al. 2009). On the other hand, it is also important to have at least 10 groups (disciplines) for each group-level predictor in order to detect Level-2 effects in a multilevel analysis (Goldstein 2011; Raudenbush et al. 2002). Since this research has four discipline-level variables (the metadata construct was removed at the later stage), it is necessary to have 40 groups to provide enough statistical power for detecting Level 2 effects. Therefore, in this research, I decided to include any disciplines which have at least 15 qualified scientists for the multilevel analysis in order to increase the number of disciplines, and five additional disciplines were included for the final sample selection (43 disciplines in total).

This research excludes (1) scientists who are 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 they do not have any data to share (155, 6.28%). In terms of the number of scientists in each discipline, this research excludes any disciplines which have less than 15 qualified scientists (304, 12.31%) or which are categorized as "others" (e.g. bioscience-other) (149, 6.03%). This results in 1,317 usable responses for the final data analysis for hypothesis testing, and out of 2,470 initial usable responses, 1,153 responses are excluded. The detailed list of the excluded respondents is indicated in Table 7.4.

Stage	Category	Frequency	Percentage
Initial Sample	Initial Responses Received	2,674	100.00%
	Not Usable Responses	204	7.63%
	Usable Responses	2,470	92.37%
Hold-Out Sample	Non-Academic Institutions	298	12.06%
	Degree and Position Requirement	247	10.00%
	No Publication (last two years)	155	6.28%
	Other Disciplines (9 disciplines & missing)	149	6.03%
	Less Than 15 Observations (41 disciplines)	304	12.31%
Final Usable Responses		1,317	53.32%

Table 7.4 Detailed List of the Excluded Respondents

7.1.3. Data Cleaning and Preparation

Data Cleaning

Data cleaning was conducted prior to the actual data analysis. The data cleaning process identifies any problems with the final field survey data to make sure that the results of data analysis are valid (Levy 2006). According to Levy (2006), there are four important reasons for data cleaning: (1) accuracy of the data collected, (2) missing data, (3) outliers, and (4) response-set. The survey data collected was reviewed by these four criteria.

First, data accuracy is important for a valid data analysis. Since this research used online survey method, errors in data collection and entry into statistics software were reduced. The final field survey data on the Web server (SurveyGizmo website) was directly transferred into an SPSS file. After the original data file was imported into SPSS, each response case was carefully reviewed about its discipline and other data collected. Some of the survey participants actually wrote their discipline names rather than choosing from the listed categories, so those data were recoded into each discipline category.

Second, each survey submission was inspected for completeness and missing values. Missing values arise when participants do not provide their answers on any item(s) or when error(s) occur in the data-collection procedure (Levy 2006). Each survey response recorded on the server was imported into SPSS with missing values. Both “Don’t Know” and “Not Applicable” responses were treated as missing values. The preliminary analysis of the original survey data shows that there are 3.05% of missing values, including user and system missing values. In the original survey data, the portions of missing values for each construct ranges from 0.47% (Perceived Career Risk) to 5.39% (Regulative Pressure by Journals), except Metadata (15.59%). The construct of metadata was found to have a large portion of missing values (2.71% of “Don’t Know,” 12.0% of “Not Applicable”, and 0.89% of system missing). One possible reason for this result would be that the questionnaire for metadata was not clear, so some of the survey participants did not interpret the questionnaire for metadata properly. The portions of missing values for each construct in both original survey data and computed score data are given in Table 7.5.

Construct	Portion of Missing Values for Each Construct	
	Original Survey Data	Computed Score Data
Funding Agencies' Regulative Pressure	3.91%	2.20%
Journals' Regulative Pressure	5.39%	3.19%
Normative Pressure	1.58%	0.84%
Metadata	15.59%	14.88%
Data Repository	3.52%	2.89%
Perceived Career Benefit	1.21%	0.38%
Perceived Career Risk	0.47%	0.15%
Perceived Effort	1.23%	0.53%
Scholarly Altruism	1.15%	0.00%
Data Sharing Behavior	1.09%	0.00%
Total	3.05%	2.20%

Table 7.5 Portion of Missing Values in Original Survey Data and Computed Score Data

This research calculates a mean score for each independent variable, and those mean scores are used for the final data analysis. This procedure somewhat reduces the portion of missing values. In the computed mean score data, the portions of missing values for each construct ranges from zero (0.0%), for Scholarly Altruism and Data Sharing Behavior), to 3.19% (Regulative Pressure by Journals), except for Metadata (which was at 14.88%). Those missing values are treated as pairwise deletion in reliability analysis, factor analysis, and ANOVA to avoid decreasing sample size and to utilize the cases with missing values. In the multilevel analysis, the missing values (for individual level variables) are treated as listwise deletion, since the multilevel regression analysis does not allow any missing values in its data analysis, and the portion of missing data in this research were small (Goldstein 2011; Heck et al. 1999; Raudenbush et al. 2002). It was found that only 12 cases out of 1,317 responses were removed in the multilevel analysis based on the listwise deletion. Since the discipline level variables were aggregated by individual responses, there was no missing value for the aggregated discipline level variables.

Third, outliers are examined in the survey data collected. Outliers refer to cases with unusual values on variables which distort statistics (Levy 2006; Tabachnick et al. 2000). The effect of outliers in this research is marginal since this research employs a large sample size for data analysis. This research employed Mahalanobis distance analysis and Cook's D to detect any outliers. Mahalanobis distance analysis utilizes the distance between a case and the mean of the remaining cases, and the value(s) more than 25 needs to be carefully examined for extreme cases (Tabachnick et al. 2000). Cook's D is used to detect outliers by measuring the effect of a case in a research model, and the value(s) more than 1 need to be investigated for unusual cases (Tabachnick et al. 2000). In this research, 22 cases were identified as possible outliers based on Mahalanobis distance analysis, but according to Cook's D, there was no case identified as outliers. Those 22 cases were carefully examined, but they were not removed for the final analysis since they have reasonable scores for each variable.

Fourth, response-sets are also examined in the survey data collected. Response-set occurs when respondents provide the same answers for all the items in a survey questionnaire (Levy 2006). The response-set problem can be detected by using a response-set test. In this research, the survey data were examined in regards to response-set, but no visible response-set was detected.

Data Preparation

When data cleaning and screening were completed, the raw items for each variable were aggregated into one composite score. A mean score was computed for each independent variable by averaging the individual item scores for each construct when there are more than two-thirds individual scores recorded (Hair et al. 2006). The data sharing behavior construct (dependent variable) was calculated by choosing the maximum frequency of data sharing behavior in five different types of data sharing behaviors (Hwang et al. 2009) since scientists' data sharing methods vary across disciplines. This yielded ten new scores for the nine independent variables and one dependent variable. The group mean was also used for the aggregated variable scale for each disciplinary level independent variable (Mayer et al. 2007).

7.2. Demographics of the Respondents

In this section, descriptive statistics for the survey participants are presented. 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 7.6 below.

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%

Demographic Category		Number	Percentage
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 7.6 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 Table 7.7.

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%

Main Discipline	Sub Discipline	Frequency	Percentage
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%

Table 7.7 Disciplines of Survey Participants

7.3. Scale Assessment

7.3.1. Construct Reliability Analysis

This section presents the scale assessments in terms of reliability and validity of the measurement items. As stated earlier, the reliability of constructs was assessed by using Cronbach's alpha indicator, which is the most common measure of scale reliability (Field 2009). Cronbach's alpha is 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. Since this study uses various survey items by other scholars, it is important to examine that the combination of the items for each construct is still valid and reliable. Cronbach's alpha values of .70 or greater are considered acceptable for the internal consistency of a construct (Hair et al. 2006; Nunnally et al. 1994). In social science, Cronbach's alpha value of .80 or more is considered more than enough, and Cronbach's alpha value of .60 is considered to be acceptable in exploratory research (Nunnally et al. 1994).

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
Metadata	3	.925	1087	.805 - .880
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 7.8 Reliability Values (N=1,317)

All of the Cronbach's alpha values for the constructs studied in this research are indicated in Table 7.8. 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. The descriptive statistics for each item are provided in the Appendix 8.11.

Each set of multiple measurement items for a construct was examined using item-to-total correlations to identify items which have measurement errors or do not share the core values of each construct. The items with low item-to-total correlation scores indicates that they do not belong to the same domain of construct and do need to be removed to increase the reliability of the measurement items for a construct (Nunnally et al. 1994). In this research, all the items have item-to-total correlations ranging from .592 to .880, which are above .50 (Field 2009). Cronbach's alpha coefficients and item-to-total correlations are indicated in Table 7.8, saying that all the research constructs have satisfactory reliability values.

7.3.2. Construct Validity Analysis

The construct validity of the measurement items was assessed by using factor analysis. The main objectives of this factor analysis are: (1) to test for convergent and discriminant validity of constructs and their relevant items and (2) evaluate the reliability of the measurement items used. In this research, principal component factor analysis with Varimax rotation was performed by extracting factors with Eigen values greater than 1. The results of factor analysis show the existence of nine factors with Eigen values greater than 1, and good convergent and discriminant validity. All of the nine observed factors explained 79.00% of the total variance, which is considered satisfactory (Hair et al. 2006). All items are loaded with factor loading value of .619 or more on each intended construct for which they were used to operationalize, showing good convergent validity. There are no cross-construct loadings above .285 for each factor, showing good discriminant validity. The factor loading value of .40 is considered as a minimum loading vale for acceptable construct validity (Field 2009; Gefen et al. 2000; Hair et al. 2006).

Convergent and discriminant validity can be ensured when a set of items for each construct load significantly (i.e. factor loading value of greater than .40) on only one factor and exhibit lower loadings (i.e. factor loading value of less than .40) on the other factors (Field 2009; Hair et al. 2006). The results of factor analysis are indicated in Table 7.9 in that a set of items measuring each construct are clustered with high factor loadings to represent a single factor.

Factors	Items	Factor Loading								
		1	2	3	4	5	6	7	8	9
Regulative Pressure by Funding Agencies	RPFA1	.227	.206	.054	-.004	**.796**	-.076	.103	.120	.076
	RPFA2	-.030	.148	.101	-.010	**.758**	.053	.168	.121	.095
	RPFA3	.181	.243	.081	.011	**.820**	-.041	.131	.093	.093
	RPFA4	.054	.279	.094	.061	**.752**	-.049	.051	.049	.055
Regulative Pressure by Journal Publishers	RPJP1	.150	**.834**	.053	-.019	.209	-.003	.131	.141	.124
	RPJP2	.007	**.808**	.125	.002	.226	-.010	.147	.075	.101
	RPJP3	.138	**.852**	.089	-.009	.213	-.015	.127	.139	.137
	RPJP4	.044	**.804**	.091	.009	.222	-.038	.082	.048	.108
Normative Pressure by Disciplines	NPD1	.264	.277	.160	.000	.279	-.076	**.619**	.184	.132
	NPD2	.187	.203	.120	-.037	.175	-.046	**.757**	.079	.230
	NPD3	.195	.053	.151	-.021	.035	-.104	**.833**	.073	.144
	NPD4	.208	.132	.139	-.067	.161	-.090	**.778**	.198	.156
Metadata	MD1	.054	.150	.058	-.045	.099	-.021	.220	.244	**.853**
	MD2	.089	.147	.055	-.034	.130	-.056	.205	.261	**.859**
	MD3	.082	.187	.094	-.047	.100	-.050	.147	.259	**.824**
Data Repository	DR1	.146	.133	.070	-.086	.133	-.074	.155	**.832**	.269
	DR2	.169	.122	.021	-.067	.156	-.068	.148	**.856**	.256
	DR3	.157	.157	.072	-.084	.122	-.066	.131	**.825**	.273
Perceived Career Benefit	PCB1	.132	.045	**.817**	-.006	.148	-.098	.065	.006	.065
	PCB2	.232	.117	**.885**	-.035	.086	-.077	.115	.032	.047
	PCB3	.285	.104	**.842**	.007	.063	-.080	.151	.046	.058
	PCB4	.265	.110	**.828**	-.046	.038	-.144	.158	.093	.038
Perceived Career Risk	PCR1	-.184	.008	-.102	.086	-.018	**.836**	-.060	.036	-.049
	PCR2	-.150	.039	-.047	.075	.000	**.887**	-.019	-.037	-.043
	PCR3	-.055	-.127	-.104	.283	-.019	**.667**	-.064	-.115	-.002
	PCR4	-.317	.001	-.146	.186	-.082	**.756**	-.140	-.103	-.029
Perceived Effort	PE1	-.138	-.009	.009	**.830**	.019	.168	-.043	-.041	-.052
	PE2	.054	.046	.019	**.881**	.023	.070	.001	.006	-.004
	PE3	-.226	-.049	-.032	**.781**	-.026	.192	-.042	-.105	-.042
	PE4	-.012	-.001	-.062	**.867**	.030	.092	-.018	-.060	-.019
Scholarly Altruism	SA1	**.786**	.138	.163	-.081	.052	-.184	.218	.195	.031
	SA2	**.819**	.111	.166	-.050	.063	-.206	.173	.165	.002
	SA3	**.794**	.075	.174	-.043	.075	-.238	.176	.174	-.005
	SA4	**.857**	.033	.162	-.058	.121	-.104	.106	.030	.094
	SA5	**.885**	.041	.183	-.081	.101	-.071	.096	.033	.088
	SA6	**.881**	.043	.183	-.074	.076	-.060	.087	.012	.077
Eigenvalue		11.10	4.32	3.16	2.31	1.99	1.73	1.44	1.39	1.02
Variance Explained		30.83%	11.99%	8.76%	6.41%	5.51%	4.81%	4.00%	3.86%	2.84%
Cumulative Variance		30.83%	42.81%	51.58%	57.99%	63.50%	68.31%	72.31%	76.17%	79.00%

Table 7.9 Results of Principal Component Factor Analysis with Varimax Rotation

7.3.3. Multi-Trait-Multi-Method (MTMM)

The validity of the instrument including convergent and discriminant validity was tested by using Multi-Trait-Multi-Method (MTMM). The MTMM calculates the correlations between each item and the other items which comprise the constructs in a study. The items comprising the same construct have high correlations (convergent validity), and the items comprising different constructs need to have low correlations (discriminant validity).

In this research, the results of MTMM show that a set of items measuring the same construct have relatively high correlations, indicating convergent validity, and the items measuring different constructs have relatively low correlations, indicating discriminant validity (except metadata and data repository). The inter-item correlation coefficients between items of the same construct ranged from .537 (Regulative Pressure by Funding Agencies) to .956 (Scholarly Altruism), which are greater than the inter-item correlation coefficients between items of the different constructs (ranged from .014. to .516) except metadata and data repository. The inter-item correlation coefficients between items of metadata and data repository range from .485 to .569, which may cause a possible multicollinearity problem. The Inter-Item and Intra-Item Correlation Matrix is included in the Appendix 8.11.

7.4. Data Preparation for Multilevel Analysis

7.4.1. Data Aggregation

More than half of hypotheses in this research are investigating the influences of the independent variables at a disciplinary (group) level (e.g. normative pressure from each discipline) on a dependent variable at an individual level (i.e. scientist's data sharing behavior). 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 et al. 2000). Therefore, it is important to check whether the aggregations of individual scientists' responses to the discipline-level variables are appropriate. In this research, Intraclass Correlations Coefficients (ICCs) including ICC(1) and ICC(2) and r_{wg} statistics (Bliese 2000; James et al. 1993) were utilized to assess whether the discipline-level variables (i.e. regulative pressures from funding agencies and journal publishers, normative pressure, metadata, and data repository) can be aggregated to the group level of analysis (Kozlowski et al. 2000).

ICC(1)

The intraclass correlation coefficients including ICC(1) and ICC(2) were examined to assess the appropriateness of data aggregation to the group level. Both ICC(1) and ICC(2) evaluate the consensus of responses within a group (Bliese 2000; Kozlowski et al. 2000); however, ICC(1) assesses the between-group variance by calculating the proportion of between-group variance to total variance that is explained by group membership (James 1982), and ICC(2) only assesses the within-group agreement in each group for the group-level variables. The following formula presents the calculation of ICC(1) value (Raudenbush et al. 2002):

$$\text{ICC(1)} = \tau_{00} / (\tau_{00}+\sigma^2)$$

Where τ_{00} is the between-group variance and σ^2 is the within-group variance.

In this research, ICC(1) measures the variances both within and between discipline(s) to assess whether scientists in the same discipline answered more similarly than did scientists across disciplines on discipline-level variables. ICC(1) can show how much discipline-level predictors vary across different disciplines. The ICC(1) values ranging from .05 to .20 are considered reasonable for between-group variance and appropriate for multilevel analysis by aggregating individual responses to the group level (Bliese 2000). In organizational studies, a minimum value of .05 for ICC(1) is considered acceptable and recommended for data aggregation and multilevel analysis (James 1982).

One-way random effect ANOVA models for each of the five disciplinary variables were conducted, and Table 7.10 below presents the within-group variance, the between-group variance, and the ICC(1) for each discipline-level measure. The 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) except metadata (.049), which is slightly below the expected range. All *p*-values for between-group variance are statistically significant, indicating that there is a significant between-discipline variance. The ICC(1) values represent that between 4.9% (i.e. metadata) to 18.2% (i.e. regulative pressure by journal publishers) of the variation in these five discipline-level variables can be explained by discipline membership. The data aggregation statistics including ICC(1) and ICC(2) are indicated in Table 7.10.

Variable	Within-Group Variance (σ^2)	Between-Group Variance (τ_{00})	ICC(1)	ICC(2)
Regulative Pressure by Funding Agencies	0.189^{***}	2.426^{**}	0.072	0.705
Regulative Pressure by Journals	0.606^{***}	2.719^{***}	0.182	0.872
Normative Pressure	0.196^{***}	2.084^{**}	0.086	0.742
Metadata	0.124^{***}	2.379^{**}	0.049	0.614
Data Repository	0.445^{***}	2.411^{***}	0.156	0.850

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

Table 7.10 Data Aggregation Statistics for Discipline-Level Variables

ICC(2)

ICC(2) was utilized to assess the reliability and validity of discipline-level measures calculated by individual scientists' responses in each discipline (Bliese 2000). In order to assess the reliability of the discipline-level means for each discipline-level construct, ICC(2) values are computed by the between-group variance and with-group variance in the ANOVA results (Dixon et al. 2006). ICC(2) values equal to or greater than .70 are considered acceptable for data aggregation (Lindell et al. 1999; Richardson et al. 2005). ICC(2) is calculated by the Spearman-Brown formula as validated by Shrout and Fleiss (Shrout et al. 1979):

$$ICC(2) = k*ICC(1) / (1+(k-1)*ICC(1))$$

Where ICC(1) is the intraclass correlation coefficient from the one-way random effect ANOVA results, and k is the average number of scientists in a discipline.

In this research, ICC(2) values for each of the discipline-level predictors were computed based on ICC(1) and the average group size (30.63). ICC(2) values are influenced by the average group size (Bliese 2000), so it is necessary to have more respondents in order to increase the reliability of group means on the discipline-level predictors. Since the average group members in each discipline is 30.63, so the ICC(2) values are rather low. ICC(2) values for regulative pressure by funding agencies (.705), regulative pressure by journal publishers (.872), normative pressure (.742), and repository (.850) were satisfactory (Table 7.10); however, ICC(2) value for metadata (.614) did not meet the cutoff criteria of .70 (Lindell et al. 1999; Richardson et al. 2005), and it means that the aggregated value for metadata has poor reliability as a discipline-level predictor.

r_{wg}

In organizational research, within-group agreement is calculated for a certain measure by each group using the r_{wg} statistic (James et al. 1993). The r_{wg} compares the observed variance on a variable in each group to the expected variance due to random error (LeBreton et al. 2008). The r_{wg} index is calculated by the following formula:

$$r_{wg} = 1 - (S_x^2 / \sigma_E^2)$$

Where S_x^2 is the observed variance on the variable X, and σ_E^2 is the expected level of random variance due to random error (James et al. 1993). In this research, the expected variance was derived from a uniform distribution of 7 point Likert scale (LeBreton et al. 2008).

In this research, multi-item indices, $r_{wg(j)}$, were employed by calculating the mean of a set of items' r_{wg} based on the above equation (James et al. 1993; Lindell et al. 1999). If there is a strong within-group agreement, the r_{wg} will become 1; if there is a strong within-group disagreement, the r_{wg} will become 0 (LeBreton et al. 2008). The r_{wg} is independent of the between-group variance level since it assesses the within-group variance for a variable only (James et al. 1993). The median r_{wg} value equal to or greater than .70 for a set of groups is recommended (Bliese 2000; James et al. 1993), suggesting sufficient within-group agreement on a group level variable. However, the median r_{wg} value of .60 is considered to be acceptable depending on research contexts (LeBreton et al. 2008).

The $r_{wg(j)}$ (within-group agreement) results are provided in Table 7.11 as an assessment on whether scientists within each discipline share similar discipline-level values (e.g. normative pressure). 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, 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. However, the median $r_{wg(j)}$ value for metadata (.54) was below the .70 recommended value and the .60 acceptable value (Bliese 2000; James et al. 1993; LeBreton et al. 2008). A full list of within-group agreement results for each discipline-level variable by each discipline was presented in the Appendix 8.13.

Group-Level Variable	r_{wg} Results			
	Average	Median	Minimum	Maximum
Regulative Pressure by Funding Agency	0.65	0.67	0.28	0.88
Regulative Pressure by Journal	0.65	0.65	0.23	0.90
Normative Pressure	0.74	0.76	0.46	0.95
Metadata	0.53	0.54	0.19	0.75
Repository	0.69	0.70	0.28	0.95

Table 7.11 Group Reliability & Within-Group Agreement Results (k=43)

The decision about data aggregation needs to be made collectively based on ICC(1), ICC(2), and $r_{wg(j)}$ (LeBreton et al. 2008). By considering any relevant indicators, four discipline-level variables (including regulative pressure by funding agency, regulative pressure by journal, normative pressure, and data repository) were aggregated to group level from the individual scientists' responses in each discipline. All the relevant indicators (i.e. ICC(1), ICC(2), and $r_{wg(j)}$) for data aggregation were within the acceptable ranges for those four discipline-level predictors.

However, one discipline-level variable, metadata, was not aggregated to group-level. The ICC(1) for metadata (.049) is slightly below the cut-off value (.05) (Bliese 2000), and the ICC(2) (.614) and the median $r_{wg(j)}$ value (.54) are below the recommended value (.70) (Lindell et al. 1999; Richardson et al. 2005) (LeBreton et al. 2008). In addition, the construct of metadata was found to have a large portion of missing values (15.59%), based on the original survey data collected. These statistical indicators for metadata show that metadata construct failed to measure the concept of metadata, and it cannot work as a discipline-level variable by data aggregation. Therefore, the metadata construct was removed from the subsequent analysis in this research.

7.4.2. Evaluation of Multilevel Regression Assumptions

A set of assumptions for multilevel regression analysis were reviewed prior to data analysis. The violations of assumptions for multilevel analysis can lead to bias of statistical analysis conducted. Regression analysis requires the following assumptions, including normality, multicollinearity, linearity, and homoscedasticity (Goldstein 2011; Heck et al. 1999; Tabachnick et al. 2000). These assumptions were tested prior to performing the multilevel analysis. The diagnoses of normality, multicollinearity, linearity, and homoscedasticity show that the main assumptions of regression analysis are not violated.

Normality

Normality of the error term is one of the key assumptions in multilevel analysis. The violation of the normality assumption can lead to bias of statistical analysis. The normality of the error term can be assessed by using visual inspections of a histogram and a normal probability plot of the standardized residuals. The histogram and the normal probability plot of the standardized residuals below (Figure 7.2 and 7.3) show that the normality of the error term was not violated.

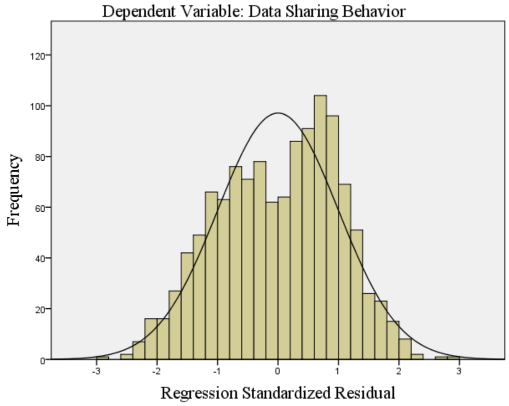


Figure 7.2 Histogram of the Standardized Residuals

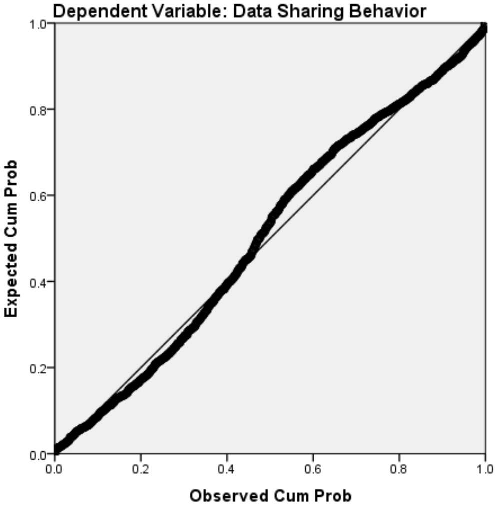


Figure 7.3 Normal Probability Plot of the Standardized Residual

Also, skewness and kurtosis are usually assessed as the measures for normality. Skewness refers to the lopsidedness of a distribution, while kurtosis refers to the peakedness or the flatness of a distribution. As kurtosis becomes close to zero, a distribution becomes normal shape; a positive value means a peaked distribution; and a negative one, a flatter distribution. According to Kline (2005), the cut-off value for extreme is 3 for skewness and 10 for kurtosis. In this research, aggregated mean scales based on a set of items for each construct were used, so normality of the error term was assessed by examining the skewness and kurtosis of the aggregated mean scales. No variable was found to have an extreme value regarding normality measures: skewness (-1.735 to 0.097) and kurtosis (-1.364 to 4.281). The Kurtosis data in Table 7.12 below shows that the distribution for scholarly altruism is slightly peaked; however, it is within the range of normality. Each variable has a normal distribution, so data transformation required for normal statistical analysis was not necessary.

	Variable	Skewness	Kurtosis
Discipline Level Predictors	Regulative Pressure by Funding Agencies	-.728	-.284
	Regulative Pressure by Journals	.097	-1.138
	Normative Pressure by Disciplines	-.656	-.339
	Data Repository	-.655	-.547
Individual Level Predictors	Perceived Career Benefit	-.487	-.345
	Perceived Career Risk	-.147	-.459
	Perceived Effort	-.419	-.134
	Scholarly Altruism	-1.735	4.281
Dependent Variable	Data Sharing Behavior	-.218	-1.364

Table 7.12 Measures of Kurtosis and Skewness for Variables

Multicollinearity

Multicollinearity is one of the important assumptions in multilevel analysis. Multicollinearity occurs when two or more independent variables are highly correlated (.80 and above) with each other in a research model (Kline 2005; Tabachnick et al. 2000). If there are high correlations among the independent variables, they cannot measure distinctive dimensions, but measure the same dimension(s) (Kline 2005). Multicollinearity distorts the data analysis of multilevel analysis by providing unstable parameter estimates. Multicollinearity can be detected by examining the Variance Inflation Factor (VIF) or the tolerance (1/VIF). If the VIF measure is greater than 10 or tolerance is less than 0.1, it raises a concern of multicollinearity (Field 2009; Hair et al. 2006). Or, strictly VIF more than 2.5 or tolerance less than .40 causes multicollinearity (Allison 1999). The presence of multicollinearity can be also examined by inspecting correlation coefficients among the independent variables in the correlation matrix. If the correlation coefficients of any two independent variables are greater than .80, this causes a concern of multicollinearity (Kline 2005; Tabachnick et al. 2000).

Variable		Not Aggregated Group-Level Variable		Aggregated Group-Level Variable	
		Toleranc e	VIF	Toleranc e	VIF
Discipline Level Predictors	Regulative Pressure by Funding	.648	1.543	.352	2.838
	Regulative Pressure by Journal	.633	1.579	.468	2.135
	Normative Pressure	.565	1.771	.441	2.267
	Data Repository	.731	1.368	.443	2.259
Individual Level Predictors	Perceived Career Benefit	.717	1.395	.756	1.322
	Perceived Career Risk	.716	1.397	.712	1.405
	Perceived Effort	.850	1.176	.837	1.195
	Scholarly Altruism	.570	1.755	.652	1.533

Table 7.13 Collinearity Statistics (DV: Data Sharing Behavior)

In this research, multicollinearity was examined by investigating VIF and correlation matrix. First, VIF was examined by running multiple regressions. The VIFs for the independent variables when data sharing behavior was treated as dependent variable are indicated in Table 7.13. As shown in Table 7.13, all VIFs of the independent variables are less than 10, showing no presence of multicollinearity (Kline 2005). Second, multicollinearity was also examined by inspecting the association between the independent variables in the correlation matrix. The correlation matrix generated by all the independent variables in this research model is presented in Table 7.14. All the correlations are less than .537, which is lower than the cut-off value of .80 for multicollinearity (Kline 2005; Tabachnick et al. 2000).

	Regulative Pressure by Funding Agency	Regulative Pressure by Journal	Normative Pressure	Repository	Perceived Career Benefit	Perceived Career Risk	Perceived Effort	Scholarly Altruism
Regulative Pressure by Funding Agency	1							
Regulative Pressure by Journal	.537**	1						
Normative Pressure	.437**	.441**	1					
Repository	.321**	.363**	.448**	1				
Perceived Career Benefit	.237**	.275**	.411**	.233**	1			
Perceived Career Risk	-.110**	-.120**	-.281**	-.217**	-.292**	1		
Perceived Effort	-.029	-.076**	-.146**	-.189**	-.165**	.381**	1	
Scholarly Altruism	.304**	.294**	.519**	.360**	.471**	-.441**	-.238**	1

Table 7.14 Correlation Matrix (with Non-Aggregated Group Level Variable)

The correlations of the aggregated group-level independent variables are indicated in Table 7.15. The correlation matrix shows slightly high statistically significant correlations among the aggregated group-level independent variables, ranging from .365 to .681; however, these correlation coefficients are lower than the cut-off value of .80 for multicollinearity (Kline 2005; Tabachnick et al. 2000).

	Regulative Pressure by Funding Agency	Regulative Pressure by Journal	Normative Pressure	Data Repository
Regulative Pressure by Funding Agency	1			
Regulative Pressure by Journal	.681**	1		
Normative Pressure	.622**	.539**	1	
Data Repository	.586**	.365**	.612**	1

Table 7.15 Correlation Matrix (Aggregated Group Level Variable Only)

Linearity

One of the important assumptions of multilevel analysis is linearity, which means that the dependent variable has a linear relationship with its independent variables. Linearity among observed variables can be assessed by inspecting scatter plots of independent and dependent variables (Tabachnick et al. 2000). Another way of detecting non-linearity is to examine residual plots, which were drawn by the standardized residuals and the standardized predicted value (Tabachnick et al. 2000). In this research, linearity assumption was ensured by observing the scatter plots based on each independent variable. The scatter plots matrix of the relationships between independent variables and dependent variable are shown in Figure 7.4. The scatter plots matrix suggests that each relationship between an independent variable and the dependent variable has a linear association.

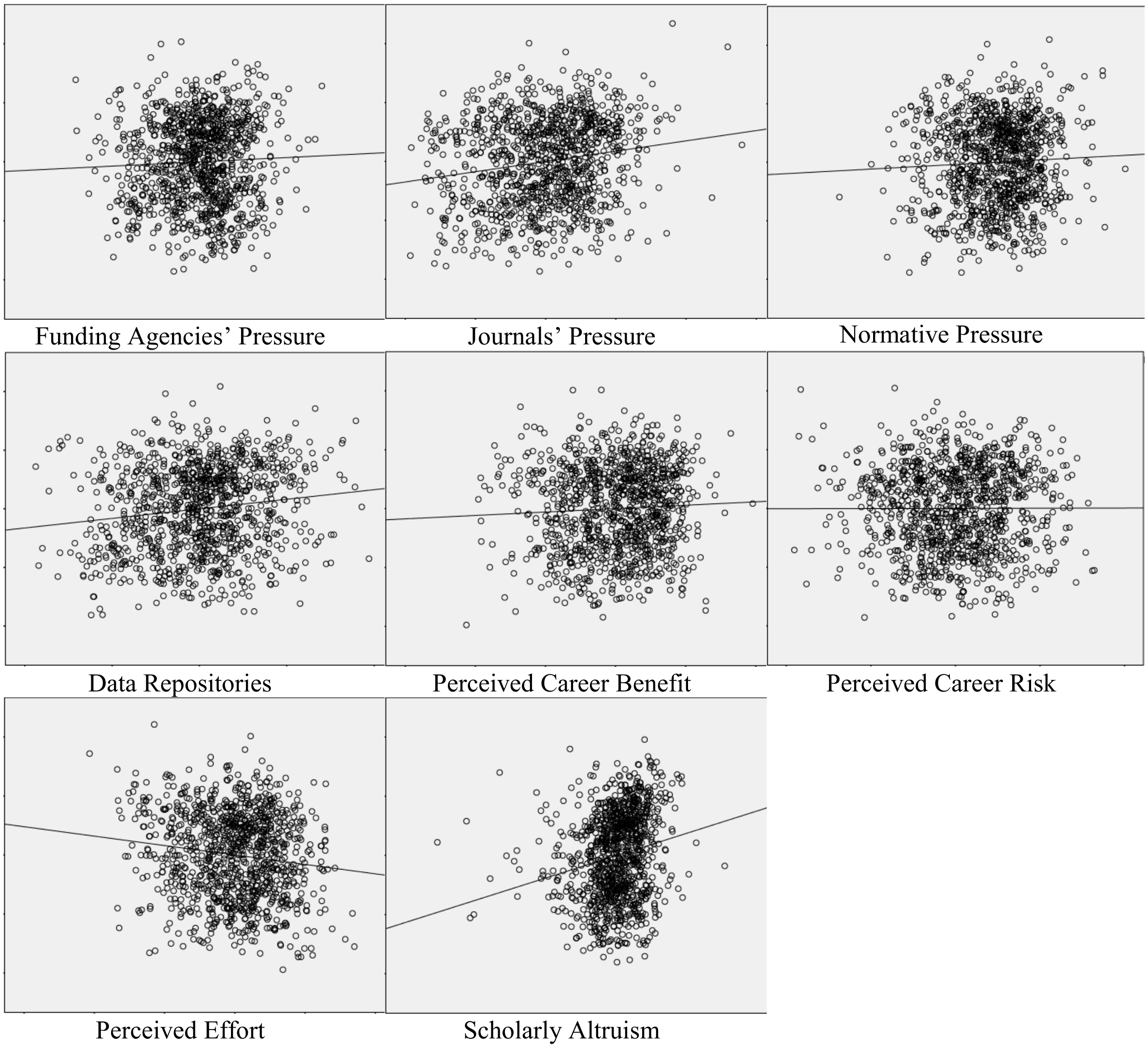


Figure 7.4 Scatter Plots Matrix

Homoscedasticity

Homoscedasticity refers to the assumption that the variance around the regression line needs to be fairly constant and same for all values of an independent variable. Homoscedasticity can cause the serious distortion of regression analysis and result (Berry et al. 1985). The homoscedasticity assumption can be assessed by visual inspection of a scatterplot of the standardized residuals and the standardized predicted values (Field 2009). Homoscedasticity can be ensured if the scatterplot of the residuals and the dependent variable shows a rectangular shape not curved or skewed shapes respectively (Tabachnick et al. 2000).

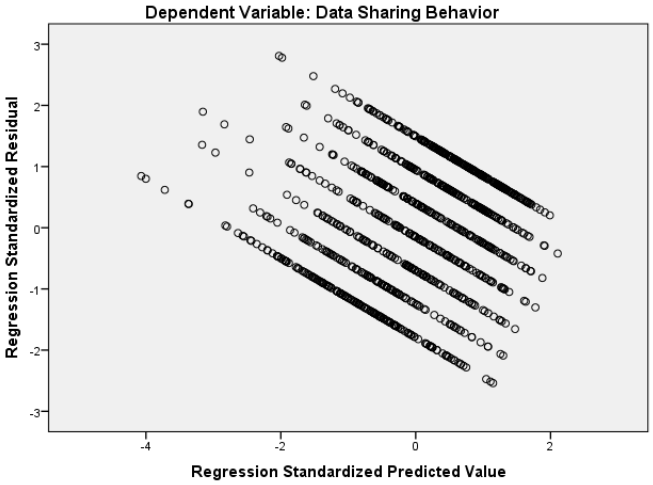


Figure 7.5 Scatterplot of Residuals and the Dependent Variable

A scatterplot of residuals and the dependent variable (Figure 7.5), shows that the scatterplot has a balanced distribution of residuals by the predicted value, and it is not curved nor skewed. Although the perfect shape for homoscedasticity is a balanced rectangular shape, the present research has a slightly rotated rectangular form. According to Fox (1991), this type of slightly rotated rectangular shape is affected by a discrete dependent variable, which is measured by n-point scales. In addition, if the sample size is large enough, the violation of homoscedasticity assumption is minimal on its data analysis and interpretation (Howell 2012).

7.5. Descriptive Statistics

7.5.1. Construct Descriptive Statistics

The scores from the multiple measurement items for each independent variable were averaged to provide an overall score for each of eight independent variables by each scientist. Then, the four discipline-level independent variables were calculated by aggregating a set of individual scientists' responses in each discipline toward each discipline-level variable. The four individual-level independent variables were the same as the average scores from the multiple items for each individual-level independent variable by each scientist. Lastly, the dependent variable was computed based on the maximum score of the diverse types of data sharing behaviors (e.g. depositing data into data repositories or providing data upon requests). The multilevel regression analysis was conducted by using these newly developed scores for each variable. The descriptive statistics of each variable were calculated including mean and standard deviation (Table 7.16).

Research Constructs		Number of Cases Used	Mean	Standard Deviation
Discipline Level Predictors	Regulative Pressure by Funding Agencies	43	4.54	0.51
	Regulative Pressure by Journal Publishers	43	3.39	0.84
	Normative Pressure	43	4.88	0.53
	Data Repository	43	4.79	0.73
Individual Level Predictors	Perceived Career Benefit	1,312	4.64	1.56
	Perceived Career Risk	1,315	4.20	1.44
	Perceived Effort	1,310	4.57	1.34
	Scholarly Altruism	1,317	6.08	1.03
Dependent Variable	Data Sharing Behavior	1,317	4.30	2.18

Table 7.16 Descriptive Statistics (N=1,317 and k=43)

7.5.2. Nonresponse Analysis

This research pays attention to nonresponse bias. Survey nonresponse can be defined as "the discrepancy between the group approached to complete a survey and those who eventually provide data" (Burkell 2003, p. 241). This study utilized several steps to reduce the nonresponse bias, based on the response facilitation approaches by Rogelberg and Stanton (2007) and Burkell (2003). The steps were: (1) make instructions clear and easy to follow; (2) present survey questions in a logical order and so they are easy to understand; (3) minimize the length of the survey and reduce the time required to be spent on the survey (i.e. 5-7 minutes); (4) provide potential survey participants with a pre-notification message in a personalized format; (5) offer a relevant incentive for the survey participants (i.e. a final report of this survey); and lastly (6) use follow-up messages in order to encourage survey participants.

This research also employs the wave analysis technique in order to detect nonresponse bias based on the data set collected (Rogelberg et al. 2007). Nonresponse analysis was conducted to check whether there are any significant differences between participating respondents and non-respondents. Babbie (1990) suggested the nonresponse analysis method which compares early responses and late responses by using the late responses as a proxy for nonresponses. In this research, the first 30% of responses were compared with the last 30% of responses to see if any significant differences existed in variables between those two groups. The first 30% of respondents participated in the survey right after the first email was sent, and the last 30% of respondents took the survey after the second and third emails (reminders) were sent. One-way Analysis of Variance (ANOVA) test was conducted on all the summated means (for independent variables) and maximum scores (for dependent variable) in order to compare the mean differences for each variable between early and late participants.

The ANOVA test shows that there are significant mean differences between the first and last groups of respondents for some of the variables including regulative pressure by funding agencies (F=4.99, p<.05), regulative pressure by journals (F=11.29, p<.01), normative pressure (F=9.21, p<.01), and data sharing behavior (F=3.91, p<.05). However, there were no significant differences between the first and last groups of respondents for the other variables including data repository (F=2.95, p=.086), perceived career benefit (F=0.31, p=.578), perceived career risk (F=0.03, p=.859), perceived risk (F=1.30, p=.255), and scholarly altruism (F=0.69, p=.406). Table 7.17 below presents the results of the ANOVA test.

Research Constructs		Round 1 (n=439)		Round 2 (n=439)		F	Sig.
		Mean	SD	Mean	SD		
Discipline Level Predictors	Regulative Pressure by Funding Agencies	4.82	1.49	4.58	1.62	4.99	.026
	Regulative Pressure by Journal Publishers	3.91	1.79	3.49	1.83	11.29	.001
	Normative Pressure by Disciplines	5.14	1.43	4.83	1.54	9.21	.002
	Data Repository	5.04	1.66	4.84	1.75	2.95	.086
Individual Level Predictors	Perceived Career Benefit	4.74	1.56	4.68	1.52	0.31	.578
	Perceived Career Risk	4.18	1.36	4.20	1.48	0.03	.859
	Perceived Effort	4.49	1.29	4.59	1.29	1.30	.255
	Scholarly Altruism	6.14	1.02	6.09	0.97	0.69	.406
Dependent Variable	Data Sharing Behavior	4.59	2.14	4.30	2.23	3.91	.048

Table 7.17 Nonresponse Analysis with Early and Late Respondents

The results of the nonresponse analysis indicate that a possible nonresponse bias exists. This is because the participants who took this survey are more likely to rate their institutional pressures (i.e. regulative pressures by funding agencies and journals and normative pressure) as high than did those who did not participate in this survey; and also because the survey participants are also more likely to share their data than non-participants. Bosnjak and colleagues (2005) also found that survey participants in Web-based surveys perceived more social pressures than non-participants. Although nonresponse biases exist in some of the variables, the effects of these nonresponses are marginal. The mean differences for each variable, divided by their averaged Standard Deviation (SD), range from 0.133 (Data Sharing Behavior) and 0.232 (Regulative Pressure by Journals), which are considered small differences (Groves 2006). In addition, following ANOVA analyses on the first and the last groups of respondents for different disciplines shows that there are no significant differences between early and late respondents in each discipline. Since the discipline-level predictors are aggregated from individual responses in each discipline, the effect of nonresponse bias by those predictors are small (canceling nonresponse biases) (Groves et al. 2008). Therefore, any weighting method for nonresponse bias was not used in this research.

7.6. Multilevel Model and Hypotheses Testing

7.6.1. Overview of Multilevel Analysis

This research employs a multilevel analysis method, which investigates 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 considers a cross-level direct-effect model, which examines how both higher-level predictors and lower-level predictors account for a lower-level outcome. In this research, the hierarchical data allows a multilevel analysis with scientists nested within their disciplines. The multilevel analysis enables 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.

A multilevel regression analysis integrates a unique random effect for each group and considers the variation of these random effects in estimating standard errors (Ethington 1997). In Ordinary Least Squares (OLS) regression, the intercept and coefficients do not vary across groups. However, in multilevel regression analysis, the intercept and coefficients are allowed to vary across groups and the variation of the intercept and coefficients are estimated. Therefore, compared to OLS regression in which individual and group level variances in a dependent variable are not estimated simultaneously, the multilevel regression analysis estimates both individual and group level residuals simultaneously (Ethington 1997; Hox 2002; Raudenbush et al. 2002).

Multilevel analysis can overcome the levels of analysis problem caused by data aggregation or disaggregation in OLS regression for multilevel data. In order to conduct OLS regression with hierarchical data, all data needs to be either disaggregated to a lower level or aggregated to a higher level. Prior studies have disaggregated the group level variables to the individual level variables by assigning each individual level unit a score representing a group level unit (Hofmann 1997); however, this violates the independence of observations assumption and causes misestimated standard errors (Ethington 1997;

Hox 2002). Raudenbush and Bryk (2002) pointed out that misestimated standard errors would increase the risk of mistakenly finding a statistically significant relationship (Type I error).

Another unit of analysis problem is the aggregation of individual level data to group level. This causes the aggregation bias, which changes the meaning of data aggregated (Ethington 1997; Hox 2002). Prior studies have aggregated the individual-level variables to the group-level variables by assigning each group-level unit a mean score for each unit. This aggregation approach loses any variance that resides at the individual level and becomes difficult to examine the cross-level relationships. Scholars have discussed the levels of analysis issues extensively (Ostroff et al. 1999), and they argued that the levels of analysis should be carefully considered in examining multilevel relationships (Dansereau et al. 1995; Sacco et al. 2003). Multilevel analysis addresses these potential problems with multilevel data by decomposing variance into different levels and integrating a unique random effect for each group (Raudenbush et al. 2002).

7.6.2. Multilevel Model

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 et al. 2002). The null model with no predictors at Level 1 and 2 was estimated as the following equations:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

Where Y_{ij} is the dependent variable (data sharing behavior for scientist i in discipline j), β_{0j} is the mean data sharing behavior in discipline j, γ_{00} is the grand mean of data sharing behavior across all disciplines, r_{ij} is within discipline variance (σ^2) in data sharing behavior, and u_{0j} is between discipline variance (τ) in data sharing behavior (Raudenbush et al. 2002). The estimation of the null model can measure the proportion of within- and between-discipline variances in the dependent variable (data sharing behavior) (Kreft et al. 1998; Wong et al. 2008). The proportion of within- and between-group variances in the null model is set to be a baseline for the changes of within- and between-group variances when both individual and group-level predictors were added into the subsequent models (Raudenbush et al. 2002).

In the second step of the multilevel modeling procedure, the within-discipline (individual level) models were created. This Level 1 model consisted of individual-level predictors including scientists' perceptions (i.e. perceived career benefit, perceived career risk, and perceived effort) and their scholarly altruism. The estimation of the Level 1 model can determine whether there was significant variance in the intercept parameters estimated at Level 1. Based on a random coefficient regression model (Hofmann 1997), the Level 1 model was estimated as the following equations:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}*(Perceived\ Career\ Benefit) + \beta_{2j}*(Perceived\ Career\ Risk)$

$+ \beta_{3j}*(Perceived\ Effort) + \beta_{4j}*(Scholarly\ Altruism) + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$ to $\beta_{4j} = \gamma_{40} + u_{4j}$

In this model, the dependent variable of scientists' data sharing behaviors (Y_{ij}) is modeled as a function of a linear combination of scientists' attitudinal perception factors and their scholarly altruism. This means that scientists' data sharing behaviors are composed of intercept β_{0j}, slopes for each discipline (β_{1j} to β_{4j}: discipline-level influences for the corresponding individual-level predictors in discipline *j*), and a random effect, r_{ij}. In this model, the intercept would vary across scientific disciplines, and the intercept is calculated by the grand mean of data sharing behavior across all disciplines, plus between random errors in data sharing behavior for each discipline. The individual-level parameters were fixed in this model, meaning that the individual level coefficients remained the same across all scientific disciplines (Raudenbush et al. 2002).

In the third step of the multilevel modeling procedure, the between discipline (group level) models were included. One of important objectives in this research is to test a set of discipline-level predictors influencing the between-discipline variance in scientists' data sharing behaviors. This multilevel model allows the total variation in scientists' data sharing behaviors to be divided into its within-discipline and between-discipline variance components. The multilevel model with predictors at Level 1 and 2 was estimated as the following equations:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}*(Perceived\ Career\ Benefit) + \beta_{2j}*(Perceived\ Career\ Risk)$

$+ \beta_{3j}*(Perceived\ Effort) + \beta_{4j}*(Scholarly\ Altruism) + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}*(Regulative\ Pressure\ by\ Funding\ Agencies) +$
$\gamma_{02}*(Regulative\ Pressure\ by\ Journal\ Publishers) + \gamma_{03}*(Normative\ Pressure) +$
$\gamma_{04}*(Data\ Repository) + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$ to $\beta_{4j} = \gamma_{40} + u_{4j}$

In this model, the intercept β_{0j} is calculated as a function of the grand mean across all disciplines on scientists' data sharing behaviors, a combination of discipline level predictors, and random error (u_{0j}) of each discipline. The γ_{01} to γ_{04} represents the influences of the corresponding discipline level predictors on scientists' data sharing behaviors. This model allows the intercept β_{0j} to vary according to discipline level predictors in a discipline. So, this model with Level 1 and Level 2 predictors can determine whether variance in the intercept β_{0j} can be explained by the Level 2 predictors. The effect u_{0j} assumes that discipline intercepts can have random errors.

7.6.3. Hypotheses Testing

Unconditional (Null) 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 et al. 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 χ^2 test for the portion of variance in data sharing behaviors between disciplines (the Level 2 residual variance of the intercept, u_0 *or* τ_{00}) was also significant (χ^2=352.065, p<.001) – between discipline variance is significantly different from zero for scientists' data sharing behaviors as a dependent variable. 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 7.18.

Fixed Effect	Coefficient	Standard Error	t-Ratio	P-Value
Data Sharing Behavior (γ_{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 7.18 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 (see Table 7.18). The Intraclass Correlation Coefficient (ICC) was calculated by the portion of disciplinary-level variance (τ_{00}) of the total variance, including disciplinary- and individual-level variances ($\tau_{00}+\sigma^2$) in the dependent variable (i.e. data sharing behavior) (Raudenbush et al. 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 Level 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 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 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 7.19 below shows the results of the individual-level model.

Predictors		Step 1	Step 2	Step 3
		Null Model	Individual-Level Predictors Only	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 (τ_{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 (σ^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 7.19 Fixed-Effect Results for Data Sharing Behavior

Multilevel Model (Individual and Discipline Level 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 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). Overall R^2 based on both discipline and individual level predictors was 0.298. The results of multilevel model including unstandardized beta and standard error, standardized beta, t-value, and p-value are shown in Table 7.20. In this research, the standardized β was not interpreted due to a possible risk to misunderstand this value since discipline-level variables are aggregated from individual responses in each group with marginal reliability (Goldstein 2011). The results of hypotheses testing are also provided below:

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 7.20 Results from Research Model (2 Level Model)

Hypothesis 1: Perceived Career Benefit

Perceived career benefit was found to have a significant, positive effect on scientists' data sharing behaviors (β=0.081(unstandardized), *p*<.05). This shows that scientists who perceive more career benefits involved in data sharing are more likely to share their data with others. Therefore, the hypothesis that "the perceived career benefit involved in data sharing positively influences scientist's data sharing behavior" was supported.

Hypothesis 2: Perceived Career Risk

Perceived career risk was not found to have a significant relationship with scientists' data sharing behaviors (β=-0.008, *p*=.839). Therefore, the hypothesis that "the perceived career risk involved in data sharing negatively influences scientist's data sharing behavior" was not supported. The null hypothesis was accepted because the large significance level shows that the result could be due to random chance.

Hypothesis 3: Perceived Effort

Perceived effort was proven to have a significant negative influence on scientists' data sharing behaviors (β=-0.138, *p*<.001). This indicates that scientists who perceive more efforts involved in data sharing are less likely to share their data with others. Therefore, the hypothesis that "the perceived effort required to share data negatively influences scientist's data sharing behavior" was supported.

Hypothesis 4: Scholarly Altruism

Scholarly altruism was found to have a significant, positive effect on scientists' data sharing behaviors (β=0.667, *p*<.001). This shows that scientists who have more scholarly altruism are more likely to share their data with others. Therefore, the hypothesis that "scientist's scholarly altruism positively influences his/her data sharing behavior" was supported.

Hypothesis 5: Regulative Pressure by Funding Agencies

Regulative pressure by funding agencies was not found to have a significant relationship with scientists' data sharing behaviors (β=-0.051, *p*=.835). Therefore, the hypothesis that "the regulative pressure by funding agencies positively influences scientist's data sharing behavior" was not supported.

Hypothesis 6: Regulative Pressure by Journals

Regulative pressure by journal publishers was proven to have a significant, positive influence on scientists' data sharing behaviors (β=0.366, *p*<.01). This indicates that scientists experiencing higher regulative pressure from journals in their disciplines are more likely to share their data with others. Therefore, the hypothesis that "the regulative pressure by journal publishers positively influences scientist's data sharing behavior" was supported.

Hypothesis 7: Normative Pressure

Normative pressure was found to have a significant, positive effect on scientists' data sharing behaviors (β=0.762, *p*<.01). This shows that both the professionalism and the expectation from peer-scientists in a scientific community positively influence scientist's data sharing behavior. Therefore, the hypothesis that "the normative pressure in a scientific discipline positively influences scientist's data sharing behavior" was supported.

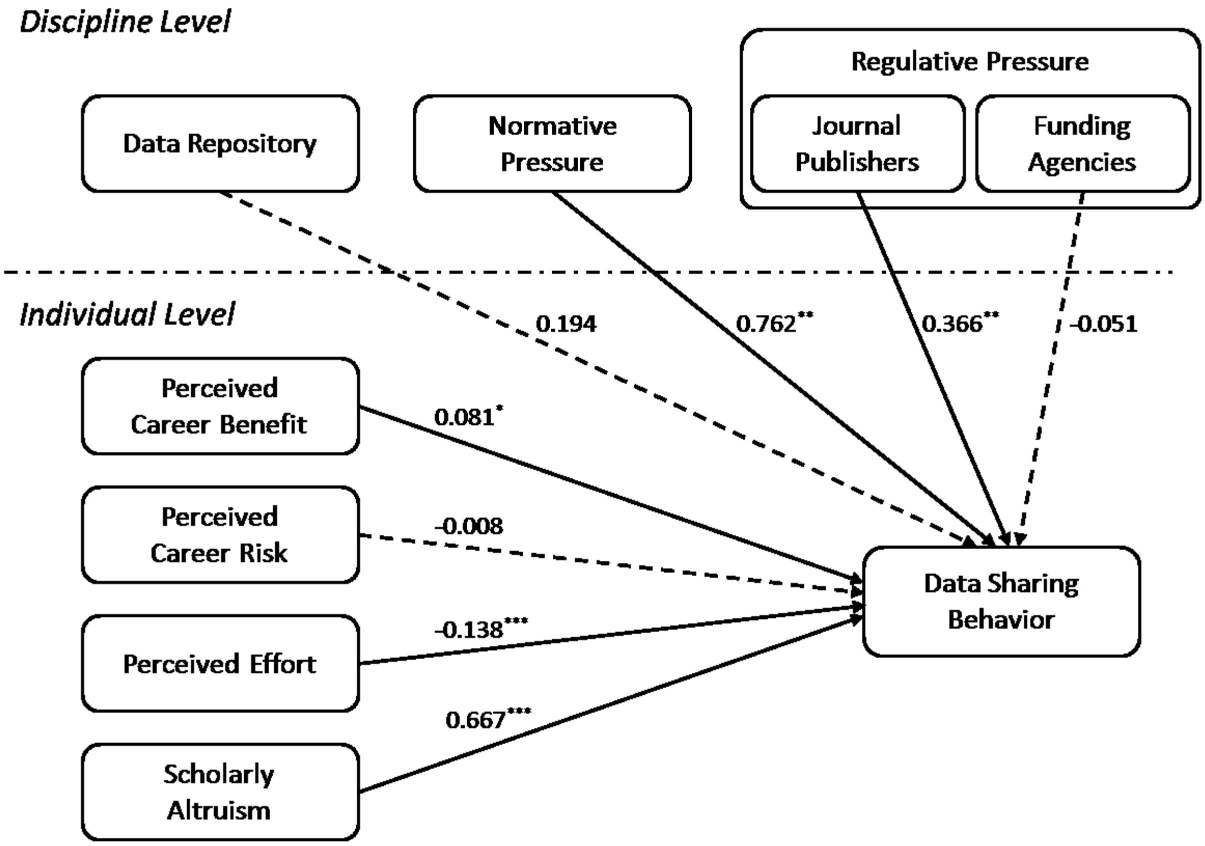
Hypothesis 8: Metadata

This hypothesis was not tested because of low internal consistency of the measurement.

Hypothesis 9: Data Repositories

The availability of data repositories in each discipline was not found to have a significant relationship with scientists' data sharing behaviors (β=0.194, *p*=.198). Therefore, the hypothesis that "the availability of data repositories in a discipline positively influences scientist's data sharing behavior" was not supported.

The summary of hypothesis testing results is shown in Figure 7.6:



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

Figure 7.6 Hypothesis Testing Results based on Scientists' Data Sharing Behavior Model

Effect Size

In this multilevel analysis, the effect sizes for the predictors which were found to be statistically significant were calculated by using Cohen's f^2 for multilevel regression (Selya et al. 2012). The following formula presents the calculation of Cohen's f^2 effect size measure (Cohen 1988):

$$f^2 = (R^2_{AB} - R^2_{A}) / (1 - R^2_{AB})$$

Where R^2_A is the proportion of variance explained by the predictor A (relative to a null model), R^2_{AB} is the proportion of variance explained by all the predictors (relative to a null model). According to Cohen's (1988) guideline for f^2 effect size measure for multilevel regression analysis, $f^2 > 0.02$ is a small effect size, $f^2 > 0.15$ is a medium effect size, and $f^2 > 0.35$ is a large effect size. Table 7.21 below shows Cohen's f^2 effect size measures for regulative pressure by journals (0.014), normative pressure (0.021), perceived career benefit (0.003), perceived effort (0.008), and scholarly altruism (0.090).

Predictor	R^2_A	R^2_{AB}	Cohen's f^2 Effect Size
Regulative Pressure by Journals	0.289	0.298	0.014
Normative Pressure	0.283		0.021
Perceived Career Benefit	0.296		0.003
Perceived Effort	0.293		0.008
Scholarly Altruism	0.235		0.090

Table 7.21 Cohen's f^2 Effect Size Measure for Multilevel Regression Analysis

7.7. Summary

The data collection procedure led to a total of 1,317 valid participants in 43 STEM disciplines for the data analysis. Those survey participants work in U.S. academic institutions, have their Ph.D. degrees, hold occupational titles of faculty, researchers, and post-docs, and currently produce research data. Once the survey data were collected, data cleaning was conducted in terms of accuracy of the data collected, missing data, outliers, and response-set. In addition, this research performed nonresponse analysis and identified a possible nonresponse bias which needs to be addressed in the interpretation of results.

Scale assessment was conducted by using Cronbach's alpha, principal component factor analysis, and multi-trait-multi-method. The results of scale assessment suggest that all the research constructs have satisfactory reliability and validity values. In addition, reliability statistics (i.e. ICC(1), ICC(2), r_{wg}) were utilized to assess the discipline-level variables can be aggregated to the group level analysis. The reliability statistics show that four discipline-level predictors can be aggregated to group level; however, metadata were found to have low reliability as a discipline-level predictor, so the metadata construct was removed from the subsequent analysis.

This research employs a multilevel analysis method in order to examine the influence of both individual and discipline level predictors on scientists' data sharing behaviors. The results of multilevel analysis show that there are significant between-discipline variances as well as within-discipline variances. At the individual level, perceived career benefit, perceived effort, and scholarly altruism were found to have significant relationships with data sharing behavior. At the institutional level, both regulative pressure by journals and normative pressure were found to have significant positive relationships with data sharing behavior.

8. Discussion and Conclusions

This chapter covers the discussions of findings and the implications of this research. Based on the results of this research, each research finding was reviewed and considered along with prior studies and their meanings. In the implications section, both theoretical and methodological contributions of this research were provided, and then practical implications were presented with regard to funding agencies, journals, professional associations, and research institutions. Lastly, the limitations of this research, and suggestions for future research were provided.

8.1. Summary of Findings

The main objective of this research is to investigate to what extent institutional and individual factors influence scientists' data sharing behaviors regarding whether they provide their data in published articles. Multilevel analysis is employed to examine both institutional and individual level effects on scientists' data sharing behaviors. Using a multilevel model, scientist's data sharing behavior is modeled at Level 1 with individual factors (i.e. perceived career benefit, perceived career risk, perceived effort, and scholarly altruism), by incorporating the institutional factors (i.e. regulative pressures by funding agencies and journals, normative pressure, and the availability of data repositories) at Level 2.

The results of multilevel analysis show that there are significant between-discipline variances (19.1%) as well as within-discipline variances (80.9%) in scientists' data sharing behaviors. 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. These four individual-level independent variables explain 16.5 percentage of the within-discipline variance (Within-Group R^2=.165).

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. These four discipline-level predictors account for 78.1 percent of the between-discipline variance in data sharing behaviors (Between-Group R^2=.781). Therefore, this research demonstrates 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).

In addition to the multilevel regression analyses, I also conducted multiple regression analyses with the same hypotheses. The multiple regression analyses show how scientists' perceptions towards institutional pressures (i.e. regulative pressures by funding agencies and journal publishers, and normative pressure) and institutional resources (i.e. metadata and data repository) along with their individual motivations (i.e. perceived career benefit and risk, perceived effort, and scholarly altruism) influence their data sharing behaviors. The results of the multiple regression analyses were presented in Table 8.1.

Fixed Effect	Unstandardized Coefficients		Standardized Beta	t-ratio	*p*-value
	Beta	Std. Error			
Discipline Level					
Funding Agencies' Pressure	0.061	0.043	0.045	1.425	0.155
Journals' Pressure	0.218	0.038	0.183	5.698	<0.001
Normative Pressure	0.081	0.051	0.056	1.578	0.115
Metadata	-0.012	0.047	-0.009	-0.261	0.794
Data Repository	0.204	0.044	0.160	4.622	<0.001
Individual Level					
Perceived Career Benefit	0.072	0.044	0.052	1.662	0.097
Perceived Career Risk	0.004	0.046	0.002	0.077	0.939
Perceived Effort	-0.200	0.046	-0.123	-4.380	<0.001
Scholarly Altruism	0.472	0.071	0.226	6.605	<0.001

Table 8.1 Results of Multiple Regression Analyses

The results of the multiple regression analyses are slightly different from the results of the multilevel regression analyses. Regulative pressure by journal publishers (β=0.218, *p*<0.001), data repository (β=0.204, *p*<0.001), perceived effort (β=-0.200, *p*<0.001), and scholarly altruism (β=0.472, *p*<0.001) were found to have significant relationships with data sharing behaviors in the multiple regression analysis. However, it was also found that regulative pressure by funding agencies (β=0.061, *p*=0.155), normative pressure (β=0.081, *p*=0.115), metadata (β=-0.012, *p*=0.794), perceived career benefit (β=0.072, *p*=0.097), and perceived career risk (β=0.004, *p*=0.939) did not have significant relationships with data sharing behaviors in the multiple regression analysis. The results of multiple regression analyses were also discussed along with the results of the multilevel analyses in the following discussion section.

8.2. Discussion of Findings

8.2.1. 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 et al. 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.

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 et al. 2008; Jones et al. 1997; Kim et al. 2009). The concept of reward through recognition and reputation is a well-known factor influencing knowledge sharing behavior (Hung et al. 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 et al. 2001; Savage et al. 2009; Stanley et al. 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. The analysis of preliminary interviews also shows that the efforts required for data sharing prevent scientists from sharing their data across different disciplines. 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 et al. 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 et al. 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 et al. 2005; Kim 2007); in the context of knowledge sharing, altruism was extensively studied and found to have significant influence on knowledge sharing (Constant et al. 1996; Davenport et al. 1998; He et al. 2009; Hung et al. 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 et al. 2010; Hung et al. 2011a; Hung et al. 2011b; Lee et al. 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 et al. 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.

8.2.2. 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 et al. 2008a); 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 et al. 2008b); 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 et al. 2008a; 2008b; Stanley et al. 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 insignificant result is that since the data sharing policy by NSF was implemented recently (National Science Foundation 2010), the effects of funding agencies' push was not reflected in scientists' data sharing behaviors as yet. The analysis of preliminary interviews shows that there are 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 NSF policy does 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 et al. 2008b; Piwowar et al. 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 which require their authors to share data by depositing it into data repositories (Savage et al. 2009; Weber et al. 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).

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 which 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 the analysis of preliminary interviews showed that the lack of data repositories was an important barrier for data sharing in several disciplines, 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 et al. 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 shows that the availability of data repositories does not necessarily increases 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.

8.3. Implications of the Study

8.3.1. Theoretical Implications

This section on theoretical implications addresses how the research findings of this study contribute to theories employed in this research. This study developed a multilevel theoretical framework by combining institutional theory and theory of planned behavior. The results of this research show that the multilevel theoretical framework proposed nicely accounts for the phenomena of scientists' data sharing. These findings have several theoretical implications for institutional theory and theory of planned behavior.

First, this research proposes a multilevel theoretical framework to investigate both institutional and individual influences on scientists' data sharing behaviors. The multilevel theoretical framework shows that scientists' data sharing behaviors are driven by individual motivations, based on their perceptions toward data sharing, along with institutional pressures in their disciplines. Although scholars have studied the diverse perceptions of scientists on their data sharing behaviors, prior studies did not fully incorporate the institutional context which also determines data sharing behaviors. Based on the multilevel theoretical framework, this research shows that both discipline-level factors (i.e. regulative pressure by journals and normative pressure in disciplines) and individual-level factors (i.e. perceived career benefit, perceived effort, and scholarly altruism) have significant influences on scientists' data sharing behaviors. The research framework integrating institutional theory and theory of planned behavior can help us understand similar social phenomena (e.g. scientists in scientific communities).

Second, with regards to institutional theory, this study sheds light on how institutional environments can influence individuals' behaviors. The results of this research show the micro-foundations of institutions by looking at institutional influences and individual motivations together. This research can advance the neo-institutional theory by applying it to the individual levels. Prior studies using institutional theory mainly focused on macro-level analysis rather than micro-level analysis, so individual actors are received less attention in the prior studies of institutional theory (Rupidara et al. 2011; Szyliowicz et al. 2010). This research examines a micro-level view of institutional theory focusing on the institutional influences (i.e. discipline level predictors) on individual scientists' behaviors as well as their motivations (i.e. individual level predictors). The results of this research show that individual scientists are influenced by institutional forces including regulative pressure by journals and normative pressure in disciplines in order to have the legitimacy of their behaviors. These findings confirm the arguments of the micro-level view of institutional theory (Carney et al. 2009; Kisfalvi et al. 2011; Mezias et al. 1994; Sitkin et al. 2005) – that social actors are influenced by institutional pressures to conform to the shared notions of appropriate behaviors (Burt 1987).

Third, the findings of this research also provide several implications for the theory of planned behavior. This research shows that individuals' perceptions can have direct influences on actual behaviors, not necessarily aggregated by attitude or mediated by intention to conduct the behavior. The results of this research show 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 et al. 2008; Watson et al. 2006; Wu et al. 2009). This research also considers actual behavior as an outcome variable, without looking at the intention to conduct a behavior. Prior studies employing the theory of planned behaviors were criticized because they did not examine actual behaviors (Ajzen 2002). This research, however, tried to measure scientists' actual data sharing behaviors with diverse means, and it was found that the measurement of actual behavior can work as an important outcome variable in the theory of planned behavior.

8.3.2. Methodological Implications

This methodological implication section covers how the research methods used in this research contribute to methodological development in the field of information science. This research has several methodological implications including (1) mixed-method approach combining qualitative and quantitative methods, (2) a multilevel regression analysis used for hierarchical data, and (3) scale development procedure taken to validate existing items and create new items for research constructs.

First, this research employed a mixed-method approach combining qualitative and quantitative approaches, and this mixed-method approach provided more fruitful outcomes in studying scientists' data sharing behaviors. The qualitative approach helped to identify diverse institutional- and individual-level predictors influencing scientists' data sharing behaviors with a rich and comprehensive context of the phenomena. The qualitative approach also assisted in the development of the research model and the design of survey. The quantitative approach helped to validate the research model by using a survey method. The quantitative approach effectively explained the phenomenon of scientists' data sharing behaviors across diverse disciplines with more generalizable results. Many scholars emphasize the synergy of using the qualitative and quantitative approaches as not opposing one another, but rather being complimentary (Creswell 2008; Greene et al. 1989; Plano Clark et al. 2008). The combination of qualitative and quantitative approaches allowed me to triangulate the research questions extensively in order to more clearly understand the phenomena of scientists' data sharing behaviors.

Second, this research utilized a multilevel analysis method by incorporating both discipline- and individual- levels to understand scientists' data sharing behaviors across diverse 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. Scholars indicated that institutional theory can operate at multiple levels, so a multilevel analysis is necessary to understand the social phenomena where each individual level is nested and interconnected with an institutional level (Oliver 1997; Thornton et al. 2008). In the integrated theoretical framework, institutional theory can account for both institutional factors and individual-level behavior, and the theory of planned behavior can explain individual-level factors and behavior. By taking a multilevel analysis method, this research showed that institutional pressures and individual motivations were closely associated with individual scientists' data sharing behaviors across different disciplines.

Third, 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. The scale development procedure followed the prescribed set of steps including item creation, scale development, and instrument testing, as proposed by Moore and Benbasat (1991). Through the scale development procedure, this research developed a set of measurement scales for the constructs studied in this research by validating existing items and creating new items. Those measurement items can be used to measure the same or similar research constructs in future research.

8.3.3. Practical Implications

This research provides several practical implications based on the results of the survey and the content analysis of preliminary interviews. This research suggests that both institutional and individual factors need to be considered in order to encourage scientists' data sharing behaviors. This section presents practical implications with regards to institutional level factors (i.e. funding agencies, journals, norms in scientific disciplines, and data repositories) and individual level factors (i.e. perceived career benefit and risk, perceived effort, and scholarly altruism).

Funding Agencies

This research suggests that funding agencies need to enforce their data sharing policies after awarding grants. The results of this research shows that regulative pressure currently exhibited by funding agencies does not have a significant effect on scientists' data sharing behaviors across different disciplines. The national funding agencies (e.g. NSF and NIH) have required their grantees to share data generated by their funds (National Institutes of Health 2003; National Science Foundation 2010). Many scientists are already aware of the data sharing policies by funding agencies. However, it is questionable whether funding agencies' data sharing policies actually exert coercive pressure on scientists' data sharing behaviors. Some scientists commented that funding agencies do not explicitly enforce their data sharing policies except in particular disciplines, so scientists do not perceive any serious coercive pressures from those policies. Therefore, in order to encourage data sharing, funding agencies need to develop a mechanism to check whether their grantees share data, and this mechanism can display more coercive pressures on scientists' data sharing behaviors.

Journals

This research shows that journals can play a critical role for encouraging scientific data sharing. Regulative pressure by journals was found to have a significant positive influence on scientists' data sharing behaviors, and this result demonstrates that journals in some disciplines exert strong coercive pressures on scientists' data sharing. Since those journals usually require data sharing as a condition of publication, scientists should observe data sharing policies by those journals. Therefore, in order to encourage data sharing, journals in each discipline need to require their authors to share data for their published articles. This can be done via two different methods: (1) mandating authors to submit their data to public repositories prior to publication (if there are any relevant repositories available in their disciplines), or (2) requiring authors with "explicit journal policy" to provide their data to those who request the data or relevant information (if there is no relevant repository available in their disciplines).

Norms in Disciplines

This research suggests that in order to facilitate data sharing, it is important to build community norms of data sharing in each scientific discipline. The results of this research show that each discipline has different norms about data sharing, and the normative pressures from each discipline significantly affect scientists' data sharing behaviors across different disciplines. Therefore, having positive normative pressures is important to support data sharing in each discipline. The normative pressures would influence scientists' data sharing behaviors in terms of social and moral obligations (Scott 2001). Education and training in each discipline can help scientists develop similar disciplinary norms about data sharing in the form of scientific ethics (DiMaggio et al. 1983), and professional associations and accreditation agencies in scientific communities can actually exert normative pressures with regards to data sharing (Grewal et al. 2002). Each scientific community can develop their norms of data sharing through education and training that are supported by their professional associations and accreditation agencies.

Data Repositories

This research shows that the availability of data repositories in each discipline does not necessarily increase scientists' data sharing behaviors across different disciplines. Although scholars argued the importance of data repositories with regards to data sharing (Brown 2003; Cragin et al. 2010; Marcial et al. 2010), we need to approach the issue of data repository carefully. It would be true that data repositories can support data sharing in certain domains of research (Marcial et al. 2010), and that the lack of data repositories can discourage scientists from sharing their data (Cragin et al. 2010), However, the correlation between the availability of data repositories and scientists' data sharing behaviors across different disciplines is doubtful. This may be caused by the fact that existing data repositories do not support scientists' data sharing (e.g. due to suitability and accessibility problems); and that the availability of data repositories alone does not encourage scientists to share their data. Therefore, scientific communities need to develop their data repositories by considering other factors (e.g. accessibility and policy guidance) to support data sharing through their data repositories.

Perceived Career Benefit and Risk

This research suggests that the scientific community needs to support scientists' receipt of more career benefits (e.g. credits and reputation) through data sharing. This research found that the perceived career benefit has a significant positive relationship with scientists' data sharing behaviors; while the perceived career risk was not found to have a significant influence on sharing the data of published articles. This means that we can encourage scientists' data sharing behaviors by providing more career benefits, rather than reducing career risks involved in data sharing. This research, however, captured the situation that in some disciplines, the current credit mechanism is not supportive of data sharing. A scholar in ecology mentioned:

> "I think that more researchers would share data if there was some way that they could be cited similarly as publications. […] Unfortunately, there is no such system, so researchers have to publish in order to improve their academic standing and have no real incentive to share data."

Therefore, the scientific community needs to provide appropriate benefits for the scientists who originally generate the data sets. In some disciplines, the academic credit mechanism needs to be adjusted to facilitate their scientists' data sharing.

Perceived Effort

This research proposes that scientific communities need to consider how to reduce scientists' efforts involved in data sharing. This research found that the perceived effort has a significant negative influence on scientists' data sharing behaviors across different disciplines. Many scientists feel that data sharing requires a significant amount of time and effort compared to a lack of rewards or incentives for sharing data. The result of this research suggests that in order to encourage data sharing, scientific communities should support scientists by helping them to organize and arrange their data sets, thus allowing the data sets to be shared with other scientists. Each scientific community can develop standardized data sharing protocols and procedures to minimize the efforts involved in sharing unstructured data sets. In addition, scientists may need institutional support for doing this, including data curation and management, which can reduce the efforts scientists need to expend in data sharing. Scientists do not have the expertise and systems to manage and curate data sets, so it would be necessary that information professionals help scientists by providing data stewardships for scientists.

Scholarly Altruism

Lastly, this research shows that scholarly altruism can support scientific data sharing. Scholarly altruism was found to have a significant positive effect on scientists' data sharing behaviors across different disciplines. This result suggests that scientists are willing to help other scientists, and to contribute to their scientific communities without expecting anything in direct return. Scholarly altruism motivates scientists to share their data with others, even though there is a lack of incentive and a significant amount of effort involved in data sharing. With the existence of scholarly altruism in scientific communities, scientists may feel grateful when other scientists share their data, and they eventually will want to reciprocate other scientists' efforts. A scholar in biology mentioned, "I find it fulfilling and stimulating to be able to hand off data I have collected (even if unpublished) to younger colleagues. The ability to look at data with new eyes and new ideas is the essence of science." Therefore, in order to facilitate data sharing, it is more important to create an altruistic culture of data sharing in scientific communities. This altruistic culture would come from the nature of scientific research, and scientific communities need to preserve this culture as an important value of science.

8.4. Limitations of the Study

This research has tried to address any possible limitations involved in its research processes; however, it has several limitations in survey instrument, data collection and analysis. In this section, I addressed the limitations of this research: (1) generalized survey instrument, (2) self-selection bias, (3) self-report problem, (4) discipline-level construct measurement, (5) deletion of metadata construct, (6) measurement problems in metadata and data repository constructs, (7) limitation of sampling strategy, and (6) small group size for several disciplines included in final analysis.

First of all, one of the main limitations is that the survey in this research did not consider domain-specific data sharing, but looked at general data sharing of published articles in diverse disciplines. Although the field survey was polished by eight subject matter experts in different disciplines through scale development process, some participants in the same discipline might approach some of the survey questions differently. For example, in certain disciplines the raw data of published articles may include materials (e.g. reagent, genetically modified organisms), specified experiment protocols, and source codes. Some participants would perceive that those are a part of their raw data associated with their published articles, but other participants in the same disciplines might not consider them in the same way. In addition, the same discipline may have different data sharing requirements and expectations depending on the types of data. The survey in this research, however, did not capture the domain-specific data sharing behaviors in various disciplines. This research focused more on general data sharing behaviors regarding the data of published articles in diverse scientific disciplines. Future research needs to investigate domain-specific data sharing behaviors.

Second, the survey method employed in this research may have self-selection bias. Although the sampling frame was randomly selected from the CoS scholar database, the field survey ultimately involved the participants who voluntarily participated in the survey. The overall response rate is only 15.28%, so the survey research may have the self-selection bias problem. This research performed a nonresponse bias test by comparing early and late respondents on each construct and found that there are significant differences in institutional pressures (i.e. regulative pressures by funding agencies and journals and normative pressure) and data sharing behaviors between those two response groups; and no significant differences in individual level predictors (i.e. perceived career benefit, perceived career risk, perceived effort, and scholarly altruism). Since the effects of those nonresponses are marginal, and the discipline level predictors are aggregated from individual responses in each discipline, the influence of nonresponse bias by those predictors might be small for this research (Groves et al. 2008). However, it is still possible that those who participated in the survey would be different from those who did not participate in terms of their data sharing behaviors. Therefore, it is necessary in future research to validate this research model with a large group of participants.

Third, another methodological limitation of survey is the self-report nature of the dependent measures. The survey method required self-report regarding the measurement of scientists' data sharing behaviors. Each participant was asked to provide their own data sharing behaviors themselves, rather than objectively observing their actual behaviors. It is impractical to examine each respondent's data sharing behaviors in diverse methods through data repositories, journal supplement, and personal communications. Therefore, scientists' self-reported data sharing behaviors can be a useful proxy for their actual data sharing behaviors. Blair and Burton (1987) also pointed that self-report measurements can be considered as relative measurements of actual behaviors.

Fourth, the multilevel method utilized in this research has several 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. In many organizational studies, it is a common method to measure group-level constructs by aggregating individuals' reports on the constructs in each group (Kraut, 1996). However, this may not measure the exact status of group-level constructs, and it may cause a potential bias in group-level measurements. In this research, the intraclass correlation coefficient (ICC) for each discipline-level construct (except metadata) was satisfactory (ranging from 0.072 to 0.182 for ICC(1) and from 0.705 to 0.872 for ICC(2)); however, the internal consistency scores ($r_{wg(j)}$) for some of the discipline-level constructs (except metadata) marginally supported for data aggregation to the discipline-level constructs (median value of $r_{wg(j)}$ ranging from 0.65 to 0.76). Therefore, the scale reliability for the discipline-level constructs needs to be carefully considered in this research.

Fifth, with regards to the fourth limitation of this research, the metadata construct failed to work as a discipline-level construct. Each survey participant was asked about the availability of metadata in their disciplines by providing the definition of metadata. Since scientists were not familiar with the term of metadata, they might interpret the term of metadata differently in spite of the definition of metadata and an example provided in this survey. The intraclass correlation coefficients were not satisfactory (0.049 for ICC(1) and 0.614 for ICC(2)), and the internal consistency score ($r_{wg(j)}$) for metadata did not support for data aggregation to the discipline-level construct (median value of $r_{wg(j)}$ for metadata is .54). Therefore, the metadata construct was removed and was not considered for the further multilevel analysis.

Sixth, the metadata and data repository constructs have their limitations to measure what they are supposed to measure. The metadata and data repository constructs were developed based on the resource-facilitating condition construct (Taylor et al. 1995; Thompson et al. 1991), which is an outdated way to measure metadata and data repositories. Therefore, the survey questions used to measure the metadata and data repository constructs might cause confusion for survey participants. The limitations of the measurements in metadata and data repository would eventually affect the quality of participants' responses. Further research should develop more accurate measurement scales for metadata and data repository based on more recent literature.

Seventh, the sampling strategy has its limitation: the discrepancy between the numbers of scientists expected to participate in the survey and the actual survey participants in each discipline. The survey research planned to recruit an equal number of participants from each discipline (equal allocation method); however, the stratified sampling strategy with equal allocation method does not work well because of the inaccuracy of scientists' disciplines registered in the CoS scholar database. There are significant differences among the numbers of survey participants in some disciplines. For example, neuroscience has 73 participants; however, public administration has only 15 participants. This discrepancy may cause a possible bias in the results of individual level analysis. In order to overcome this limitation, it is necessary to use a more reliable scholar database and to recruit more people in the disciplines which have less participants compared to other disciplines.

Lastly, 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 organization studies (Hox 2002; Scherbaum et al. 2009), this research included five disciplines (out of forty-three disciplines) which contain less 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 their internal consistency; however, this research decided to include those five disciplines in order to increase the statistical power to detect the discipline-level (Level 2) predictors. 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). Excluding disciplines with fewer members would reduce statistical power, and make Level-2 estimates unstable (Type II error).

Despite those limitations, this research allows us to examine how discipline-level and individual-level predictors influence scientists' data sharing behaviors across diverse scientific disciplines. This would be the first empirical study investigating both disciplinary environments and individual motivations with regards to scientists' data sharing behaviors. Future research can improve the current research by considering the aforementioned limitations, and I provided possible directions for such future research with regards to scientists' data sharing behaviors.

8.5. Suggestions for Future Research

This section provides suggestions for future research based on the findings of this research. Future research can (1) investigate some of the research constructs employed in this research, (2) examine the discipline differences in data sharing practices, (3) compare the data sharing factors in different major disciplines, (4) consider organizational-level factors influencing scientists' data sharing behaviors, and (5) expand on the issues of data reuse along with data sharing.

First, future research in scientific data sharing should expand upon the relationships examined in this research. Future research can investigate some of the research constructs employed in this research more carefully. Contrary to the earlier arguments (McCullough et al. 2008; Piwowar et al. 2008a), regulative pressure by funding agencies was not found to have a significant relationship with data sharing behaviors. Future research can examine this construct as an individual level predictor (i.e. perception toward the regulative pressure by funding agencies) by considering individual scientists' funding sources, or it might be interesting to re-investigate this construct as a discipline-level predictor several years in the future (after the NSF grantees have had chances to share the data they collected through the support of their funding agencies). In addition, the constructs of both data repository and metadata need to be re-examined; a researcher can objectively measure each of those constructs by investigating their availabilities in each discipline. Then, those measurements can be entered as objective and accurate discipline-level data in a multilevel analysis.

Second, along with the results of this research, future research can examine how the discipline and individual level factors influencing scientists' data sharing behaviors differ across different disciplines and what factors contribute to those differences. The discipline comparison study can illustrate domain-specific data sharing behaviors, and their different patterns of discipline- and individual-level predictors that motivate and prevent scientists' data sharing behaviors. Since each discipline has its own historical, institutional, and research dependent contexts, each discipline has its own pattern of factors influencing scientists' data sharing behaviors, and the different patterns of factors can be compared among distinctive scientific disciplines. Especially, both interviews and archival study can be employed to understand the context and sequential nature of scientists' data sharing to explore the underlying meanings of their data sharing behaviors in different disciplines.

Third, with regard to the second future research direction, researchers can investigate how the discipline and individual level factors influence scientists' data sharing behaviors in different major disciplines (e.g. biological sciences or engineering). Each subordinate discipline can be aggregated into its superordinate discipline (e.g. physics under physical sciences) or categorized into one domain discipline based on their shared research interests (e.g. animal science under agricultural sciences). The hypotheses in this research can be tested with a set of relevant disciplines, which can be grouped into one superordinate or domain discipline. The results of the hypotheses testing with one set of disciplines can be compared and contrasted with another set of disciplines. This future research can illustrate how the discipline and individual level factors affect scientists' data sharing behaviors in one group of disciplines, as compared to another group of disciplines in similar and/or different ways.

Fourth, future research needs to consider organizational-level factors influencing scientists' data sharing behaviors as well as disciplinary- and individual-level factors. The current research did not address the organizational issues (e.g. organizational supports and resources involved in scientists' data sharing behaviors). Some of the survey participants commented that academic institutions influence their data sharing behaviors either negatively by concerning potential intellectual property involved in their scientists' research, or positively by supporting their scientists with organizational resources (i.e. institutional data repositories and data management supports). For future research, we need to consider organizational influences along with disciplinary and individual influences on data sharing behaviors.

Lastly, researchers also need to consider data reuse issues along with data sharing. Data sharing is not the final outcome, but reuse of data would be the final goal of data sharing. This research focuses on data sharing in the perspective of providing data; however, it is very important to understand data reuse in the perspective of actively utilizing existing data sets. Future research needs to examine how scientists locate, interpret, and understand existing data sets for their own research in view of a data reuse perspective. Also, future research can investigate the factors influencing data sharing and reuse simultaneously, and explore the relationship between scientists' data sharing and reuse behaviors.

8.6. Conclusions

This research has investigated how both institutional environments and individual motivations influence scientists' data sharing behaviors across diverse disciplines. The results of this research show 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 suggest that in order to encourage data sharing, we need to consider both institutional environments and individual motivations simultaneously.

This research has methodological and theoretical implications. A mixed-method approach was employed, including interview study, to examine what kinds of institutional and individual factors influence scientists' data sharing in diverse disciplines, and survey study, to investigate to what extent those factors influence scientists' data sharing behaviors across different disciplines. This research proposed the multilevel theoretical framework combining institutional theory and theory of planned behavior, which was found to nicely account for scientists' data sharing behaviors across diverse disciplines. Then, 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).

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 shows 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.

Therefore, future research needs to investigate how data sharing factors differ across different disciplines, and what contribute to those differences. Furthermore, 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.

Appendices

Appendix 1. Preliminary Study Interview Questions

Questions about Current Research and Data Use

- What is your research field(s) and what kinds of research do you do?
- What kinds of data do you usually generate for your research?

Questions about Data Sharing

- Would you tell me whether and how researchers (including you) in your field share their data?
- Do researchers have any data repositories, portals, and tools? What are they?
- Would you believe that you have the authority to decide whether you make some or all of your data available for the public?

Questions about Factors Influencing Data Sharing

- What motivates researchers (including you) in your field to share their data?
- What prevents researchers (including you) in your field from sharing their data?
- Would you feel that you have enough support available to you when you share your data? If not, what kind of support would you need that you are not currently getting?

Questions about the Role of Data Sharing in Scientific Research

- What would you say to the idea that data sharing is critical for novel scientific findings?
- What would you say to the idea that data sharing among researchers will improve your research performance? How would you think data sharing help you to conduct your research?

Appendix 2. Field Survey Distribution in Each Discipline

Disciplines		Registered Scientists	Initial Email	Email with Survey	Returned Email	Retired	Student	Not Scientist	Total	Opt Out
Engineering	Aerospace Engineering	2,913	394	289	2	6		1	9	6
	Agricultural Engineering	1,441	370	279		2	3		5	5
	Biomedical Engineering	8,873	377	277	2	3		1	6	8
	Chemical Engineering	4,513	391	288		4	2	3	9	11
	Civil Engineering	6,584	363	292		11	2	1	14	12
	Computer Engineering	10,441	383	288	8	3	4	1	16	12
	Electrical Engineering	10,376	372	303	6	13	2		21	6
	Environmental Eng.	7,273	383	311	2	4	3		9	6
	Industrial Engineering	2,988	376	304		2			2	5
	Mechanical Engineering	11,744	376	299		7	1		8	8
Physical Sciences	Astronomy	10,930	385	293	4	11	3		18	7
	Chemistry	17,084	376	281	5	5	2		12	7
	Physics	24,982	378	283		6	1		7	4
Earth, Atmospheric, & Ocean Sci.	Geology	10,689	387	319	6	8	2		16	7
	Marine Biology	2,572	379	313	1	4		4	9	12
	Ocean Science	4,517	386	308	1	2			3	9
Com. Sci.	Computer Science	30,680	391	319		10		4	14	3
Agricultural Sciences	Animal Science	3,324	371	287	5	2	2		9	7
	Food Science & Tech.	2,992	387	290	1	4		1	6	9
	Forestry	2,726	378	275	2	4	2	1	9	8
	Natural Resource Conservation	1,940	372	276	2	8			10	12
	Plant Pathology	1,919	358	240	2	2	1		5	1
	Wildlife & Wetlands Sci.	1,666	387	314	2	3	4		9	10
	Horticulture	2,001	368	266	3	2			5	10
Biological Sciences	Biochemistry	13,446	376	286	11	1	2	2	16	9
	Biological Science	28,313	366	290	3		3	7	13	13
	Bioinformatics	3,039	372	304	8		6	8	22	7
	Biophysics	3,783	394	283	7	4	2	2	15	4
	Botany	2,493	385	307	5	7	1	2	15	2
	Cell Biology	9,745	366	296	3	3	3		9	7
	Developmental Biology	2,068	382	305	5	10		1	16	5
	Ecology	4,006	383	315	5	6	5	1	17	13
	Entomology	2,586	390	300	2	2			4	10
	Genetics	7,301	398	286	3	6	2	3	14	13

Disciplines		Registered Scientists	Initial Email	Email with Survey	Returned Email	Retired	Student	Not Scientist	Total	Opt Out
	Microbiology	7,634	394	324	2	1	4	3	10	5
	Molecular Biology	7,710	389	319		3	3	2	8	8
	Neuroscience	14,143	376	295	3	5	3	1	12	8
	Zoology	1,273	384	312		2	1		3	17
	Biotechnology	5,580	368	303	6	6		1	13	5
Psychology	Psychology	25,677	1167	871	5	22	4	2	33	18
Social Sciences	Anthropology	9,195	381	319	6	8	3		17	14
	Geography	7,710	394	308	6	4	2		12	11
	Political Science	12,896	382	297	11	6	2		19	18
	Public Administration	9,822	376	302	6	10	1		17	15
	Sociology	12,484	393	306	10	5	5		20	16
Health Fields	Anesthesiology	10,180	335	237	4				4	2
	Dentistry	10,174	396	290	6	6			12	13
	Neurology	7,998	396	283	6	1			7	8
	Nursing	20,628	362	281		2			2	10
	Obstetrics & Gynecology	8,214	313	199	4	1			5	2
	Oncology	24,746	376	295	7		1		8	4
	Pediatrics	18,419	313	230	1				1	5
	Pharmacy	7,447	382	297	3	1			4	4
	Psychiatry	18,193	352	253	2	3			5	6
	Radiology	10,342	363	264	2	1			3	3
	Surgery	21,261	297	202	1				1	2
Total			21,789	16,753	197	252	87	52	588	462

Appendix 3. Survey Items for Pre-Test

Constructs	Items	Sources
Regulative Pressure by Funding Agencies	1. In my discipline, data sharing is mandated by public funding agencies' policy.	(Kostova et al. 2002)
	2. In my discipline, there is public funding agencies' policy to require researchers to share data.	(Liang et al. 2007)
	3. In my discipline, there are public funding agencies to promote and enforce data sharing.	(Kostova et al. 2002)
	4. In my discipline, data sharing policy by public funding agencies is strictly enforced.	(Kostova et al. 2002)
	5. In my discipline, public funding agencies force researchers to share data.	(Shi et al. 2008)
	6. In my discipline, public funding agencies can penalize researchers in some manner if they do not share data.	(Teo et al. 2003)
	7. In my discipline, if researchers do not share data, public funding agencies will punish them.	(Ke et al. 2009)
	8. In my discipline, if researchers do not share data as public funding agencies ask, something bad will happen to them.	(Ke et al. 2009)
Regulative Pressure by Journal Publisher	In my discipline, data sharing is mandated by journals' policy.	(Kostova et al. 2002)
	In my discipline, there is journals' policy to require researchers to share data.	(Liang et al. 2007)
	In my discipline, there are journals to promote and enforce data sharing.	(Kostova et al. 2002)
	In my discipline, data sharing policy by journals is strictly enforced.	(Kostova et al. 2002)
	In my discipline, journals force researchers to share data.	(Shi et al. 2008)
	In my discipline, journals can penalize researchers in some manner if they do not share data.	(Teo et al. 2003)
	In my discipline, if researchers do not share data, journals will punish them.	(Ke et al. 2009)
	In my discipline, if researchers do not share data as journals ask, something bad will happen to them.	(Ke et al. 2009)
Normative	In my discipline, it is expected that researchers	(Kostova et al. 2002)

Constructs	Items	Sources
Pressure	would share data.	
	In my discipline, data sharing is a moral obligation.	(Kostova et al. 2002)
	In my discipline, researchers care a great deal about data sharing.	(Kostova et al. 2002)
	In my discipline, researchers share their data even if not required by policies.	(Kostova et al. 2002)
	In my discipline, data sharing is at the heart of who we are as researchers.	(Kostova et al. 2002)
	In my discipline, the extent to which data sharing is adopted by my peer researchers is high.	(Liang et al. 2007)
	In my discipline, many researchers are currently participating in data sharing.	(Son et al. 2007)
	In my discipline, data sharing has been widely adopted by researchers.	(Liu et al. 2010)
Metadata	In my discipline, researchers can easily access metadata.	(Cho 2006; Thompson et al. 1991)
	In my discipline, metadata are available for researchers to share data.	(Taylor et al. 1995)
	In my discipline, there are not enough metadata to help researchers share data.	(Taylor et al. 1995)
	In my discipline, due to lack of metadata, researchers have found data sharing is difficult.	(Neufeld et al. 2007)
	In my discipline, researchers have metadata necessary to share data.	(Thompson et al. 1991; Venkatesh et al. 2003)
	In my discipline, data sharing is very supportive due to metadata.	(Cheung et al. 2000)
	In my discipline, the current metadata does not support data sharing.	(Neufeld et al. 2007)
Repository	In my discipline, researchers can easily access data repositories.	(Cho 2006; Thompson et al. 1991)
	In my discipline, data repositories are available for researchers to share data.	(Taylor et al. 1995)
	In my discipline, there are not enough data repositories to help researchers share data.	(Taylor et al. 1995)
	In my discipline, due to lack of data repositories, researchers have found data sharing is	(Neufeld et al. 2007)

Constructs	Items	Sources
	difficult.	
	In my discipline, researchers have data repositories necessary to share data.	(Thompson et al. 1991; Venkatesh et al. 2003)
	In my discipline, data sharing is very supportive due to data repositories.	(Cheung et al. 2000)
	In my discipline, the current data repositories do not support data sharing.	(Neufeld et al. 2007)
Perceived Career Benefit	I can earn academic credits such as more citations by sharing data.	(Bock et al. 2005)
	Data sharing would enhance my academic recognition.	(McLure Wasko et al. 2000)
	Data sharing would improve my status in a research community.	(McLure Wasko et al. 2000)
	Data sharing can give me a possible opportunity to collaborate with other researchers.	(Chiu et al. 2006)
	Data sharing will provide me with possible authorships.	(Chiu et al. 2006)
	Data sharing can help me to build my reputation in a research community.	(Chiu et al. 2006)
	I can earn respect from other researchers by sharing data.	(Bock et al. 2005)
	I can gain some academic rewards by sharing data.	(Bock et al. 2005)
	Data sharing would be helpful in my academic career.	<New>
	Data sharing can demonstrate the quality of my research work.	<New>
Perceived Career Risk	There is a high probability of losing publication opportunities if I share data.	(Featherman et al. 2003)
	Data sharing may cause my research ideas to be stolen by other researchers.	(Featherman et al. 2003)
	My shared data may be misused or misinterpreted by other researchers.	(Featherman et al. 2003)
	I would label data sharing as a potential loss.	(Pavlou 2003)
	I believe that overall riskiness of data sharing is high.	(Pavlou 2003)
	Sharing data may jeopardize my control over the data.	(Hu et al. 2002)
	If I share data, I may suffer loss from	(Liu et al. 2008)

Constructs	Items	Sources
	irresponsible behaviors from other researchers.	
	If I share data, I may suffer loss from opportunistic behaviors from other researchers.	(Liu et al. 2008)
Perceived Effort	Sharing data involves too much time for me (e.g. to organize/annotate).	(Thompson et al. 1991)
	Sharing data takes too much time from my normal duties.	(Thompson et al. 1991)
	I need to make a significant effort to share data.	(Davis 1989)
	It is free of effort for me to share data.	(Klein 2007)
	I would find data sharing easy to do.	(Davis et al. 1989)
	It would be easy for me to become skillful at sharing data.	(Klein 2007)
	I would find data sharing difficult to do.	(Davis et al. 1989)
	Overall, data sharing requires a significant amount of time and effort.	(Davis 1989)
Scholarly Altruism	I am willing to help other researchers by sharing data.	(Kankanhalli et al. 2005)
	I share data so that other researchers can conduct their research more easily.	(Kankanhalli et al. 2005)
	I share data so that other researchers can utilize it for their research.	<New>
	I share data so that other researchers have access to original data sets.	<New>
	I share data to support open scientific research.	<New>
	I share data to support better scientific research.	(Baytiyeh et al. 2010)
	I share data to help improve the quality of scientific research.	(Baytiyeh et al. 2010)
	By sharing data, I want to contribute to scientific development.	<New>
Data Sharing Behavior	In the last two years, how frequently do you deposit your data into disciplinary data repositories (including interdisciplinary data repositories)?	<New>
	In the last two years, how frequently do you deposit your data into institutional data repositories (provided by universities or research institutions)?	<New>
	In the last two years, how frequently do you upload data into “public” Web spaces (personally managed, non-disciplinary and	<New>

Constructs	Items	Sources
	non-institutional data repositories)?	
	In the last two years, how frequently do you provide data by publishing supplementary materials (along with your article)?	<New>
	In the last two years, how frequently do you provide your data via personal communication methods upon request?	<New>

Appendix 4. Pre-Test Analysis and Results

Regulative Pressure by Funding Agencies

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: In my discipline, data sharing is mandated by public funding agencies' policy.	3.72	.922	.608	.861
Item2: In my discipline, there is public funding agencies' policy to require researchers to share data.	3.66	1.010	.378	.884
Item3: In my discipline, there are public funding agencies to promote and enforce data sharing.	3.41	.946	.544	.867
Item4: In my discipline, data sharing policy by public funding agencies is strictly enforced.	2.55	.948	.625	.859
Item5: In my discipline, public funding agencies force researchers to share data.	2.62	1.115	.708	.850
Item6: In my discipline, public funding agencies can penalize researchers in some manner if they do not share data.	2.59	1.086	.795	.840
Item7: In my discipline, if researchers do not share data, public funding agencies will punish them.	2.34	1.010	.706	.851
Item8: In my discipline, if researchers do not share data as public funding agencies ask, something bad will happen to them.	2.41	1.086	.701	.851

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 2, 3	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 2	
Rule3: Redundant Item or Not Working well for measurement	Item 7, 8 (Similar to Item 6); Item 2 (Similar to 1)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.874
Final Items (Item 1, 4, 5, 6)	4	.809

Regulative Pressure by Journals

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: In my discipline, data sharing is mandated by journals' policy.	2.89	1.219	.831	.885
Item2: In my discipline, there is journals' policy to require researchers to share data.	2.70	1.265	.747	.895
Item3: In my discipline, there are journals to promote and enforce data sharing.	2.52	1.051	.552	.910
Item4: In my discipline, data sharing policy by journals is strictly enforced.	1.85	.770	.781	.893
Item5: In my discipline, journals force researchers to share data.	2.04	.940	.770	.891
Item6: In my discipline, journals can penalize researchers in some manner if they do not share data.	2.11	.974	.804	.888
Item7: In my discipline, if researchers do not share data, journals will punish them.	1.93	.829	.806	.890
Item8: In my discipline, if researchers do not share data as journals ask, something bad will happen to them.	1.96	.808	.450	.915

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 3, 8	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 3, 8	
Rule3: Redundant Item or Not Working well for measurement	Item 7, 8 (Similar to 6); Item 2 (Similar to 1)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.908
Final Items (Item 1, 4, 5, 6)	4	.885

Normative Pressure

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: In my discipline, it is expected that researchers would share data.	3.69	1.072	.799	.912
Item2: In my discipline, data sharing is a moral obligation.	3.52	1.122	.611	.927
Item3: In my discipline, researchers care a great deal about data sharing.	3.07	1.067	.719	.918
Item4: In my discipline, researchers share their data even if not required by policies.	3.21	1.082	.770	.914
Item5: In my discipline, data sharing is at the heart of who we are as researchers.	3.52	1.153	.716	.919
Item6: In my discipline, the extent to which data sharing is adopted by my peer researchers is high.	2.93	1.132	.787	.913
Item7: In my discipline, many researchers are currently participating in data sharing.	3.28	1.032	.810	.911
Item8: In my discipline, data sharing has been widely adopted by researchers.	3.00	1.069	.779	.913

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	None	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 2	
Rule3: Redundant Item or Not Working well for measurement	Item 6, 8 (Similar to 7); Item 5 (Not Many Studies)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.926
Final Items (Item 1, 3, 4, 7)	4	.866

Metadata

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: In my discipline, researchers can easily access metadata.	2.77	.908	.648	.812
Item2: In my discipline, metadata are available for researchers to share data.	2.96	.958	.533	.834
Item3: In my discipline, there are not enough metadata to help researchers share data.	2.73	.827	.700	.803
Item4: In my discipline, due to lack of metadata, researchers have found data sharing is difficult.	2.88	.816	.546	.828
Item5: In my discipline, researchers have metadata necessary to share data.	2.88	.711	.733	.801
Item6: In my discipline, data sharing is very supportive due to metadata.	2.73	.604	.500	.834
Item7: In my discipline, the current metadata does not support data sharing.	3.12	.653	.570	.825

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 4, 6, 7	Item 2 (Exception)
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	None	
Rule3: Redundant Item or Not Working well for measurement	Item 6, 7 (Confusing/Low Variance)	

Items Considered	Number of Items	Cronbach's α
Original Items	7	.842
Final Items (Item 1, 2, 5)	3	.820

Data Repository

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: In my discipline, researchers can easily access data repositories.	3.36	1.026	.672	.759
Item2: In my discipline, data repositories are available for researchers to share data.	3.54	.922	.535	.786
Item3: In my discipline, there are not enough data repositories to help researchers share data.	2.68	.819	.218	.835
Item4: In my discipline, due to lack of data repositories, researchers have found data sharing is difficult.	3.18	.945	.521	.789
Item5: In my discipline, researchers have data repositories necessary to share data.	2.96	.744	.798	.746
Item6: In my discipline, data sharing is very supportive due to data repositories.	2.89	.737	.470	.797
Item7: In my discipline, the current data repositories do not support data sharing.	3.39	.875	.661	.763

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 3, 4, 6	Item 2 (Exception)
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 3	
Rule3: Redundant Item or Not Working well for measurement	Item 6, 7 (Confusing)	

Items Considered	Number of Items	Cronbach's α
Original Items	7	.809
Final Items (Item 1, 2, 5)	3	.851

Perceived Career Benefit

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: I can earn academic credits such as more citations by sharing data.	2.79	1.166	.637	.908
Item2: Data sharing would enhance my academic recognition.	3.00	1.089	.830	.894
Item3: Data sharing would improve my status in a research community.	3.14	.970	.780	.898
Item4: Data sharing can give me a possible opportunity to collaborate with other researchers.	3.82	.905	.688	.904
Item5: Data sharing will provide me with possible authorships.	3.14	.891	.425	.918
Item6: Data sharing can help me to build my reputation in a research community.	3.39	.956	.811	.896
Item7: I can earn respect from other researchers by sharing data.	3.43	.836	.637	.907
Item8: I can gain some academic rewards by sharing data.	2.71	1.013	.563	.911
Item9: Data sharing would be helpful in my academic career.	3.21	.995	.801	.897
Item10: Data sharing can demonstrate the quality of my research work.	3.61	.916	.659	.905

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 5, 8	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 5	
Rule3: Redundant Item or Not Working well for measurement	Item 6, 7 (Similar to 2); 4, 10 (Less Relevant)	

Items Considered	Number of Items	Cronbach's α
Original Items	10	.913
Final Items (Item 1, 2, 3, 9)	4	.859

Perceived Career Risk

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: There is a high probability of losing publication opportunities if I share data.	2.69	.930	.669	.885
Item2: Data sharing may cause my research ideas to be stolen by other researchers.	3.07	.923	.658	.886
Item3: My shared data may be misused or misinterpreted by other researchers.	3.41	.780	.716	.880
Item4: I would label data sharing as a potential loss.	2.55	.686	.566	.893
Item5: I believe that overall riskiness of data sharing is high.	2.72	.841	.721	.879
Item6: Sharing data may jeopardize my control over the data.	3.24	.951	.645	.887
Item7: If I share data, I may suffer loss from irresponsible behaviors from other researchers.	3.07	.842	.646	.886
Item8: If I share data, I may suffer loss from opportunistic behaviors from other researchers.	3.24	.830	.833	.869

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 4	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	None	
Rule3: Redundant Item or Not Working well for measurement	Item 6, 7, 8 (Not Many Studies)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.896
Final Items (Item 1, 2, 3, 5)	4	.843

Perceived Effort

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: Sharing data involves too much time for me (e.g. to organize/annotate).	3.14	.970	.723	.891
Item2: Sharing data takes too much time from my normal duties.	3.18	.905	.883	.876
Item3: I need to make a significant effort to share data.	3.36	.870	.730	.891
Item4: It is free of effort for me to share data.	3.68	.670	.449	.912
Item5: I would find data sharing easy to do.	3.29	.810	.707	.893
Item6: It would be easy for me to become skillful at sharing data.	2.93	.900	.521	.909
Item7: I would find data sharing difficult to do.	2.79	.917	.758	.888
Item8: Overall, data sharing requires a significant amount of time and effort.	3.32	.863	.819	.882

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 4, 6	
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	Item 4, 6	
Rule3: Redundant Item or Not Working well for measurement	Item 2 (Similar to 1); 5 (Similar to 7)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.905
Final Items (Item 1, 3, 7, 8)	4	.887

Scholarly Altruism

Items Used for Pretest	Mean	SD	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
Item1: I am willing to help other researchers by sharing data.	3.83	.602	.535	.846
Item2: I share data so that other researchers can conduct their research more easily.	3.52	.738	.458	.859
Item3: I share data so that other researchers can utilize it for their research.	3.66	.670	.630	.835
Item4: I share data so that other researchers have access to original data sets.	3.62	.622	.475	.853
Item5: I share data to support open scientific research.	3.97	.566	.616	.838
Item6: I share data to support better scientific research.	3.90	.618	.769	.819
Item7: I share data to help improve the quality of scientific research.	3.79	.675	.740	.821
Item8: By sharing data, I want to contribute to scientific development.	3.97	.566	.616	.838

Rules	Item(s) Removed	Note
Rule1: Low Item-Total Correlation (<.60)	Item 4	Item 1, 2 (Exception)
Rule2: Cronbach's α if Item Deleted > Overall Cronbach's α	None	Item 2 (Exception)
Rule3: Redundant Item or Not Working well for measurement	Item 8 (Similar to 6)	

Items Considered	Number of Items	Cronbach's α
Original Items	8	.856
Final Items (Item 1, 2, 3, 5, 6, 7)	6	.831

Appendix 5. Changes in Measurement Items (after Pre-Test)

Item #	Constructs & Items	Changes Made
	Regulative Pressure by Funding Agencies	
1	In my discipline, data sharing is mandated by the policy of public funding agencies.	public funding agencies' policy > the policy of public funding agencies
2	In my discipline, data sharing policy of public funding agencies is enforced.	by > of strictly > removed
3	In my discipline, public funding agencies require researchers to share data.	force > require
4	In my discipline, public funding agencies can penalize researchers if they do not share data.	will > can
	Regulative Pressure by Journal Publisher	
1	In my discipline, data sharing is mandated by journals' policy.	No change
2	In my discipline, data sharing policy of journals is enforced.	by > of strictly > removed
3	In my discipline, journals require researchers to share data.	force > require
4	In my discipline, journals can penalize researchers if they do not share data.	will > can
	Normative Pressure	
1	In my discipline, it is expected that researchers would share data.	No change
2	In my discipline, researchers care a great deal about data sharing.	No change
3	In my discipline, researchers share data even if not required by policies.	No change
4	In my discipline, many researchers are currently participating in data sharing.	No change
	Metadata	
1	In my discipline, researchers can easily access metadata.	No change
2	In my discipline, metadata are available for researchers to share data.	No change
3	In my discipline, researchers have the metadata necessary to share data.	No change

Item #	Constructs & Items	Changes Made
	Repository	
1	In my discipline, researchers can easily access data repositories.	No change
2	In my discipline, data repositories are available for researchers to share data.	No change
3	In my discipline, researchers have the data repositories necessary to share data.	No change
	Perceived Career Benefit	
1	I can earn academic credit such as more citations by sharing data.	No change
2	Data sharing would enhance my academic recognition.	No change
3	Data sharing would improve my status in a research community.	No change
4	Data sharing would be helpful in my academic career.	No change
	Perceived Career Risk	
1	There is a high probability of losing publication opportunities if I share data.	No change
2	Data sharing may cause my research ideas to be stolen by other researchers.	No change
3	My shared data may be misused or misinterpreted by other researchers.	No change
4	I believe that the overall riskiness of data sharing is high.	No change
	Perceived Effort	
1	Sharing data involves too much time for me (e.g. to organize/annotate).	No change
2	I need to make a significant effort to share data.	No change
3	I would find data sharing difficult to do.	easy > difficult
4	Overall, data sharing requires a significant amount of time and effort.	No change
	Scholarly Altruism	
1	I am willing to help other researchers by sharing data.	No change
2	I would share data so that other researchers can conduct their research more easily.	would (added)

Item #	Constructs & Items	Changes Made
3	I would share data so that other researchers can utilize it for their research.	would (added)
4	I would share data to support open scientific research.	would (added)
5	I would share data to contribute to better scientific research.	want to > would share data to better (added)
6	I would share data to help improve the quality of scientific research.	would (added)
	Data Sharing Behavior	
1	In the last two years, how frequently have you deposited your data into disciplinary data repositories for every article?	do you deposit > have you deposited
2	In the last two years, how frequently have you deposited your data into institutional data repositories for every article?	do you deposit > have you deposited
3	In the last two years, how frequently have you uploaded your data into "public" Web spaces for every article?	do you upload > have you uploaded
4	In the last two years, how frequently have you provided access to your data by publishing supplementary materials for every article?	provide data > have you provided access to your data
5	In the last two years, how frequently have you responded to the request(s) by providing data via personal communication methods?	provide your data > have you responded to the request(s) by providing data

Appendix 6. Email Messages Used

1st Email Contact

Title: Introduction to Survey on Scientists' Data Sharing

Dear Dr. [Last Name]:

Hello, my name is Youngseek Kim, and I am a doctoral candidate in the School of Information Studies at Syracuse University. I have been studying scientists' data sharing and reuse in diverse science and engineering disciplines.

A few days from now I plan to send you an email requesting your participation in a brief online survey about scientists' data sharing behaviors. You have been randomly selected to participate in this survey from the Community of Scientists' Profile Database. I am writing you in advance because I have found that many people like to know ahead of time that they will be contacted for such activity.

The survey focus is on the experience of scientists who generate research data, and who may or may not share their data of published articles with other scientists outside their research group(s). This study is an important one that will help the stakeholders of scientific research (e.g. scientists, funding agencies, journals, and research institutions) to better understand the factors facilitating and preventing the researchers' current data sharing behaviors. This study is approved by the Institutional Review Board at Syracuse University (#IRB11-243). Please visit the project website (http://ykim58.mysite.syr.edu) to know more about the research and researcher.

As a token of appreciation for your participation, survey participants will be entered to win one of ten $50 eGift Cards. All survey participants also will receive the final report of this survey.

Thank you for your time and consideration.

Sincerely,

Youngseek Kim

Doctoral Candidate
School of Information Studies
Syracuse University

2nd Email Contact

Title: Survey on Scientists' Data Sharing

Dear Dr. [Last Name]:

Hello. I am Youngseek Kim, a doctoral candidate in the School of Information Studies at Syracuse University. A few days ago, I contacted you regarding a survey on scientists' data sharing behaviors.

I am writing to ask your help in conducting this important research, which investigates the reasons why scientists make their decisions to share or not to share the research data of published articles with other scientists outside their research group(s).

Previously, I interviewed a number of scientists in diverse science and engineering disciplines to explore domain specific data sharing practices and to investigate the factors facilitating and preventing the researchers' current data sharing behaviors. Based on my prior study, I have developed a brief survey to further investigate my prior findings and to compare the researchers' data sharing behaviors in different disciplines.

I am cordially inviting you to participate in this survey. It will take you about five to seven minutes to complete. The survey is anonymous and does not collect any identification information. You can provide a great deal of assistance by taking a few minutes to share your experiences about data sharing.

Please follow this link to reach the survey:

[Survey link]

As a token of appreciation for your participation, survey participants will be entered to win one of ten $50 eGift Cards. All survey participants also will receive the final report of this survey.

Thank you very much for considering assisting me with this important study.

Sincerely,

Youngseek Kim

Doctoral Candidate
School of Information Studies
Syracuse University

3rd Email Contact (1st Reminder)

Title: Reminder: Survey on Scientists' Data Sharing

Dear Dr. [Last Name]:

Greetings. Last month, I contacted you regarding a survey on scientists' data sharing behaviors. This note comes as a reminder to ask if you would participate in the survey.

If you have already completed and submitted the online survey, please accept my sincere thanks. If not, I would like to ask if you are able to complete the survey sometime this week. I would be especially grateful for your help, since it is only by asking researchers like you to share your experience that we can understand why scientists decide to share or not to share their research data with other scientists.

In addition, the quality of the survey will depend on the response rate, so I am depending upon you to help with this important effort. The survey will take you about five to seven minutes to complete. It is anonymous and does not collect any identification information.

Please follow this link to reach the survey:

[Survey link]

Thank you for your support.

Sincerely,

Youngseek Kim

Doctoral Candidate
School of Information Studies
Syracuse University

P.S. If you would prefer to opt out of further emails regarding this study, please reply back to this message.

4^{th} Email Contact (2^{nd} and Last Reminder)

Title: Final Reminder: Survey on Scientists' Data Sharing

Dear Dr. [Last Name]:

Hello. About two month ago, I contacted you regarding my research survey on scientists' data sharing behaviors. The survey is now drawing to a close, and this is the last contact I plan to make with the random sample of scientists who are registered in the Community of Scientists' Profile Database regarding participation.

This survey looks at the experience of scientists who generate research data and may or may not share their data of published articles with other scientists outside their research group(s). This study is an important one that will help the stakeholders of scientific research to better understand the factors facilitating and preventing the researchers' current data sharing behaviors.

I wanted to get in touch one more time since I am concerned that scientists who have not responded may have different experiences than those who have. Hearing from everyone in this small discipline-wide sample helps assure that the survey results are as accurate as possible.

Consequently, I would like to ask again for your participation in this survey. It will take about five to seven minutes to complete. This study is approved by the Institutional Review Board at Syracuse University (#IRB11-243). Please visit the project website (http://ykim58.mysite.syr.edu) to know more about the research and researcher. Please note that this survey is anonymous and does not collect any identification information.

Please follow this link if you plan to respond to the survey:

[Survey link]

Thank you for your time and consideration.

Sincerely,

Youngseek Kim

Doctoral Candidate
School of Information Studies
Syracuse University

Appendix 7. Final Survey Instrument

Thank you for your willingness to participate in this survey.

Completion of this survey is entirely voluntary. The survey is anonymous and does not collect any identification information. All answers will be reported as aggregated data. You can drop out at any time and for any reason without penalty.

In order to appreciate your participation, the following benefits will be provided for the survey participants.

> (1) Particpants who complete the survey and submit their email address will be entered to win one of ten $50 eGift Cards.
>
> (3) All survey participants also will receive the final report of this survey.

Please provide your email address at the end of this survey if you would like to be entered to win one of eGift Cards and receive the final report of this survey.

If you have any inquiries about this survey, please let me know by email (ykim58@syr.edu) or phone (315-464-0824). If you have any concerns about your rights as a participant, contact the Office of Research Integrity and Protections at Syracuse University by email (orip@syr.edu) or phone (315-443-3013).

To begin this survey, please click the NEXT button below.

By proceeding to the survey I acknowledge that I have read the above statements and that I am 18 years of age or older.

<NEXT BUTTON>

(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

***NOTE**: Metadata is a set of data that provides information about one or more aspects of the original research data (e.g. Ecological Metadata Language).

5. Metadata*

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 metadata.	1	2	3	4	5	6	7	8	9
Metadata are available for researchers to share data.	1	2	3	4	5	6	7	8	9
Researchers have the metadata necessary to share data.	1	2	3	4	5	6	7	8	9

6. Data Repositories

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

7. 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

8. 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

9. 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

10. 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

11. 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

12. 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+

13. 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
13. (Continued) In the last two years, how frequently have you…	Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Every time	Not Applicable	Do Not Know
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

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

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

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

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

18. 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)

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

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

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

Once you click "SUBMIT" button below, you will be redirected to a separate page, where you can provide your email address to be entered to a drawing and receive the final report of this survey.

Appendix 8. Response Rate by Discipline

	Discipline	Sample	Response	Rate
Engineering	Aerospace Engineering	280	21	7.50%
	Agricultural Engineering	274	24	8.76%
	Biomedical Engineering	271	31	11.44%
	Chemical Engineering	279	48	17.20%
	Civil Engineering	278	41	14.75%
	Computer Engineering	272	18	6.62%
	Electrical Engineering	282	39	13.83%
	Engineering Science and Engineering Physics	-	4	-
	Environmental Engineering	302	33	10.93%
	Industrial/Manufacturing Engineering	302	23	7.62%
	Mechanical Engineering	291	35	12.03%
	Metallurgical and Materials Engineering	-	25	-
	Nuclear Engineering	-	2	-
	Engineering, other	-	12	-
Physical Sciences	Astronomy	275	36	13.09%
	Chemistry	269	47	17.47%
	Physics	276	46	16.67%
	Physical Sciences, other	-	13	-
Earth, Atmospheric, and Ocean Sciences	Atmospheric Sciences	-	29	-
	Geosciences (Geology)	303	77	25.41%
	Ocean Sciences	305	61	20.00%
	Earth, Atmospheric, and Ocean Sciences, other (Marine Biology)	304	26	8.55%
Mathematical Sciences	Mathematics and Applied Mathematics	-	6	-
	Statistics	-	11	-
Computer Science	Computer Science	305	25	8.20%
Agricultural Sciences	Agricultural Sciences	-	44	-
	Animal Sciences	278	33	11.87%
	Fishing and Fisheries Sciences	-	21	-
	Food Sciences (Food Sciences & Technology)	284	15	5.28%
	Forestry	266	38	14.29%
	Natural Resources Conservation	266	28	10.53%
	Plant Sciences (Plant Pathology)	235	55	23.40%
	Soil Sciences	-	15	-
	Wildlife and Wildlands Management	305	19	6.23%
	Agricultural Sciences, other (Horticulture)	261	17	6.51%
Biological Sciences	Anatomy	-	4	-
	Biochemistry	270	71	26.30%
	Biology (Biological Science)	277	32	11.55%
	Biometry and Epidemiology (Bioinformatics)	282	25	8.87%
	Biophysics	268	29	10.82%
	Botany	292	25	8.56%
	Cell Biology	287	49	17.07%
	Developmental Biology	289	41	14.19%
	Ecology	298	89	29.87%
	Entomology and Parasitology	296	29	9.80%

	Discipline	Sample	Response	Rate
	Genetics	272	60	22.06%
	Microbiology, Immunology, and Virology	314	78	24.84%
	Molecular Biology	311	77	24.76%
	Neuroscience	283	80	28.27%
	Nutrition	-	14	-
	Pathology	-	7	-
	Pharmacology	-	16	-
	Physiology	-	29	-
	Zoology	309	20	6.47%
	Biosciences, other (Biotechnology)	290	14	4.83%
Psychology	Clinical Psychology	838	27	3.22%
	Psychology, Except Clinical		46	-
	Psychology, Combined		22	-
Social Sciences	Agricultural Economics	-	5	-
	Anthropology	302	38	12.58%
	Economics	-	16	-
	Geography	296	37	12.50%
	History and Philosophy of Science	-	4	-
	Linguistics	-	3	-
	Political Science	278	43	15.47%
	Public Administration	285	24	8.42%
	Sociology	286	42	14.69%
	Social Sciences, other	-	54	-
Health Fields	Anesthesiology	233	9	3.86%
	Cardiology	-	3	-
	Communication Disorders Sciences	-	1	-
	Dental Sciences (Dentistry)	278	18	6.47%
	Endocrinology	-	5	-
	Gastroenterology	-	3	-
	Hematology	-	2	-
	Neurology	276	7	2.54%
	Nursing	279	28	10.04%
	Obstetrics and Gynecology	194	5	2.58%
	Oncology/Cancer Research	287	19	6.62%
	Ophthalmology	-	4	-
	Pediatrics	229	12	5.24%
	Pharmaceutical Sciences (Pharmacy)	293	17	5.80%
	Preventive Medicine and Community Health	-	23	-
	Psychiatry	248	9	3.63%
	Pulmonary Disease	-	2	-
	Radiology	261	9	3.45%
	Surgery	201	10	4.98%
	Veterinary Sciences	-	10	-
	Clinical Medicine, other	-	14	-
	Health Related, other	-	20	-
Others	Other Disciplines		49	
	Missing		23	
	Total	16,165	2,470	15.28%

Appendix 9. Demographics of Field Survey Respondents

Demographic Category		Number	Percentage
Gender	Male	1,735	70.24%
	Female	680	27.53%
	Missing	55	2.23%
Age	under 24	7	0.28%
	25-34	349	14.13%
	35-44	576	23.32%
	45-54	576	23.32%
	55-64	613	24.82%
	65+	322	13.04%
	Missing	27	1.09%
Ethnic	Asian/Pacific Islander	352	14.25%
	Black/African-American	34	1.38%
	Caucasian	1,881	76.15%
	Hispanic	67	2.71%
	Native American/Alaska Native	9	0.36%
	Other/Multi-Racial	64	2.59%
	Missing	63	2.55%
Education	Associates Degree	2	0.08%
	Bachelors Degree	39	1.58%
	Masters Degree	202	8.18%
	PhD/Doctoral Degree	2,202	89.15%
	Missing	25	1.01%
Position	Graduate Student	148	5.99%
	Lecturer/Instructor	46	1.86%
	Professor of Practice	10	0.40%
	Post-Doctoral Fellow	147	5.95%
	Researcher	210	8.50%
	Assistant Professor	334	13.52%
	Associate Professor	491	19.88%
	Full Professor	807	32.67%
	Professor Emeritus	121	4.90%
	Other	140	5.67%
	Missing	16	0.65%
Status	Tenured	1220	49.39%
	On Tenure Track	296	11.98%
	Not On Tenure Track	737	29.84%
	Retired	138	5.59%
	Missing	79	3.20%
Sector	Academic	2,172	87.94%
	Government	157	6.36%
	Non-profit	53	2.15%
	Commercial	47	1.90%
	Other	19	0.77%
	Missing	22	0.89%
Total		2,470	100%

Appendix 10. Research Disciplines of Field Survey Respondents

Main Discipline	Sub Discipline	Frequency	Percentage
Engineering	Aerospace Engineering	21	0.85%
	Agricultural Engineering	24	0.97%
	Biomedical Engineering	31	1.26%
	Chemical Engineering	48	1.94%
	Civil Engineering	41	1.66%
	Computer Engineering	18	0.73%
	Electrical Engineering	39	1.58%
	Engineering Science and Engineering Physics	4	0.16%
	Environmental Engineering	33	1.34%
	Industrial/Manufacturing Engineering	23	0.93%
	Mechanical Engineering	35	1.42%
	Metallurgical and Materials Engineering	25	1.01%
	Nuclear Engineering	2	0.08%
	Engineering, other	12	0.49%
Physical Sciences	Astronomy	36	1.46%
	Chemistry	47	1.90%
	Physics	46	1.86%
	Physical Sciences, other	13	0.53%
Earth, Atmospheric, and Ocean Sciences	Atmospheric Sciences	29	1.17%
	Geosciences	77	3.12%
	Ocean Sciences	61	2.47%
	Earth, Atmospheric, and Ocean Sciences, other	26	1.05%
Mathematical Sciences	Mathematics and Applied Mathematics	6	0.24%
	Statistics	11	0.45%
Computer Science	Computer Science	25	1.01%
Agricultural Sciences	Agricultural Sciences	44	1.78%
	Animal Sciences	33	1.34%
	Fishing and Fisheries Sciences	21	0.85%
	Food Sciences	15	0.61%
	Forestry	38	1.54%
	Natural Resources Conservation	28	1.13%
	Plant Sciences	55	2.23%
	Soil Sciences	15	0.61%
	Wildlife and Wildlands Sciences	19	0.77%
	Agricultural Sciences, other	17	0.69%
Biological Sciences	Anatomy	4	0.16%
	Biochemistry	71	2.87%
	Biology	32	1.30%
	Biometry and Epidemiology	25	1.01%
	Biophysics	29	1.17%
	Botany	25	1.01%
	Cell Biology	49	1.98%
	Developmental Biology	41	1.66%
	Ecology	89	3.60%
	Entomology and Parasitology	29	1.17%
	Genetics	60	2.43%

Main Discipline	Sub Discipline	Frequency	Percentage
	Microbiology, Immunology, and Virology	78	3.16%
	Molecular Biology	77	3.12%
	Neuroscience	80	3.24%
	Nutrition	14	0.57%
	Pathology	7	0.28%
	Pharmacology	16	0.65%
	Physiology	29	1.17%
	Zoology	20	0.81%
	Biosciences, other	14	0.57%
Psychology	Clinical Psychology	27	1.09%
	Psychology, Except Clinical	46	1.86%
	Psychology, Combined	22	0.89%
Social Sciences	Agricultural Economics	5	0.20%
	Anthropology	38	1.54%
	Economics	16	0.65%
	Geography	37	1.50%
	History and Philosophy of Science	4	0.16%
	Linguistics	3	0.12%
	Political Science	43	1.74%
	Public Administration	24	0.97%
	Sociology	42	1.70%
	Social Sciences, other	54	2.19%
Health Fields	Anesthesiology	9	0.36%
	Cardiology	3	0.12%
	Communication Disorders Sciences	1	0.04%
	Dental Sciences	18	0.73%
	Endocrinology	5	0.20%
	Gastroenterology	3	0.12%
	Hematology	2	0.08%
	Neurology	7	0.28%
	Nursing	28	1.13%
	Obstetrics and Gynecology	5	0.20%
	Oncology/Cancer Research	19	0.77%
	Ophthalmology	4	0.16%
	Pediatrics	12	0.49%
	Pharmaceutical Sciences	17	0.69%
	Preventive Medicine and Community Health	23	0.93%
	Psychiatry	9	0.36%
	Pulmonary Disease	2	0.08%
	Radiology	9	0.36%
	Surgery	10	0.40%
	Veterinary Sciences	10	0.40%
	Clinical Medicine, other	14	0.57%
	Health Related, other	20	0.81%
	Other	49	1.98%
	Missing	23	0.93%
Total		2,470	100.00%

Appendix 11. Descriptive Statistics and Reliability for Final Survey Items

Construct	Item	Mean	SD	Number of Responses	Cronbach's alpha	Number of Cases Used
Regulative Pressure by Funding Agencies	RPFA1	5.37	1.94	1,283	.867	1,210
	RPFA2	4.01	1.81	1,258		
	RPFA3	5.17	1.96	1,270		
	RPFA4	4.00	1.86	1,251		
Regulative Pressure by Journal Publishers	RPJP1	4.06	2.23	1,274	.911	1,177
	RPJP2	3.36	1.87	1,217		
	RPJP3	3.78	2.19	1,263		
	RPJP4	3.06	1.86	1,230		
Normative Pressure by Disciplines	NPD1	5.14	1.78	1,301	.875	1,269
	NPD2	4.85	1.77	1,299		
	NPD3	4.89	1.73	1,296		
	NPD4	4.90	1.76	1,289		
Metadata	MD1	4.02	1.70	1,122	.925	1,087
	MD2	4.19	1.70	1,118		
	MD3	4.05	1.69	1,095		
Data Repository	DR1	4.93	1.83	1,277	.931	1,251
	DR2	5.10	1.77	1,277		
	DR3	4.67	1.81	1,258		
Perceived Career Benefit	PCB1	4.34	1.89	1,293	.922	1,273
	PCB2	4.71	1.71	1,308		
	PCB3	4.89	1.62	1,308		
	PCB4	4.61	1.70	1,295		
Perceived Career Risk	PCR1	4.13	1.74	1,313	.867	1,301
	PCR2	4.26	1.72	1,312		
	PCR3	4.68	1.59	1,309		
	PCR4	3.72	1.75	1,309		
Perceived Effort	PE1	4.49	1.58	1,302	.877	1,277
	PE2	4.86	1.54	1,297		
	PE3	4.02	1.60	1,302		
	PE4	4.90	1.57	1,302		
Scholarly Altruism	SA1	6.11	1.166	1,312	.948	1,256
	SA2	6.11	1.168	1,313		
	SA3	6.02	1.270	1,303		
	SA4	5.94	1.188	1,301		
	SA5	6.16	1.051	1,298		
	SA6	6.18	1.050	1,284		

(Field Study: N=1,317)

Appendix 12. Inter-Item and Intra-Item Correlation Matrix (MTMM Matrix)

		Regulative Pressure by Funding Agencies				Regulative Pressure by Journal Publishers				Normative Pressure by Disciplines				Metadata			Data Repository		
		RPFA1	RPFA2	RPFA3	RPFA4	RPJP1	RPJP2	RPJP3	RPJP4	NPD1	NPD2	NPD3	NPD4	MD1	MD2	MD3	DR1	DR2	DR3
Regulative Pressure by Funding Agencies	RPFA1	1																	
	RPFA2	.537**	1																
	RPFA3	.841**	.574**	1															
	RPFA4	.564**	.598**	.611**	1														
Regulative Pressure by Journal Publishers	RPJP1	.492**	.304**	.496**	.345**	1													
	RPJP2	.361**	.488**	.394**	.418**	.697**	1												
	RPJP3	.467**	.338**	.506**	.372**	.883**	.711**	1											
	RPJP4	.345**	.362**	.386**	.479**	.638**	.734**	.692**	1										
Normative Pressure by Disciplines	NPD1	.485**	.322**	.492**	.325**	.516**	.368**	.494**	.359**	1									
	NPD2	.354**	.348**	.369**	.274**	.389**	.366**	.395**	.288**	.644**	1								
	NPD3	.235**	.192**	.228**	.178**	.251**	.218**	.241**	.187**	.575**	.626**	1							
	NPD4	.338**	.309**	.350**	.268**	.350**	.298**	.352**	.264**	.618**	.650**	.709**	1						
Metadata	MD1	.255**	.276**	.272**	.183**	.350**	.295**	.350**	.277**	.379**	.429**	.343**	.406**	1					
	MD2	.281**	.269**	.295**	.206**	.340**	.306**	.354**	.304**	.405**	.442**	.352**	.428**	.857**	1				
	MD3	.262**	.233**	.276**	.205**	.351**	.312**	.354**	.314**	.382**	.412**	.307**	.391**	.761**	.792**	1			
Data Repository	DR1	.282**	.244**	.279**	.181**	.334**	.276**	.336**	.244**	.391**	.351**	.305**	.399**	.569**	.557**	.521**	1		
	DR2	.312**	.250**	.309**	.206**	.352**	.269**	.349**	.247**	.395**	.354**	.293**	.408**	.508**	.558**	.485**	.834**	1	
	DR3	.278**	.248**	.293**	.185**	.355**	.274**	.357**	.248**	.387**	.342**	.300**	.388**	.530**	.550**	.548**	.794**	.829**	1
Perceived Career Benefit	PCB1	.188**	.160**	.201**	.138**	.201**	.180**	.208**	.151**	.288**	.256**	.215**	.265**	.178**	.197**	.190**	.167**	.135**	.162**
	PCB2	.205**	.156**	.228**	.146**	.254**	.235**	.258**	.195**	.345**	.303**	.309**	.329**	.181**	.195**	.189**	.196**	.186**	.213**
	PCB3	.221**	.144**	.222**	.159**	.252**	.231**	.257**	.194**	.379**	.327**	.321**	.346**	.184**	.211**	.209**	.211**	.203**	.226**
	PCB4	.208**	.164**	.216**	.150**	.263**	.247**	.268**	.199**	.369**	.342**	.334**	.354**	.181**	.219**	.225**	.224**	.217**	.256**
Perceived Career Risk	PCR1	-.095**	0.024	-.073**	-0.053	-.058*	-0.028	-.073**	-.063*	-.183**	-.124**	-.206**	-.178**	-0.051	-.092**	-.071*	-.125**	-.125**	-.121**
	PCR2	-.094**	0.026	-0.038	-0.044	-0.03	-0.029	-0.03	-0.044	-.172**	-.100**	-.175**	-.150**	-0.045	-.089**	-.076*	-.149**	-.147**	-.137**
	PCR3	-.147**	-0.044	-.119**	-0.045	-.176**	-.115**	-.182**	-.151**	-.218**	-.139**	-.158**	-.189**	-.133**	-.145**	-.145**	-.207**	-.181**	-.182**
	PCR4	-.195**	-0.048	-.156**	-.107**	-.149**	-.089**	-.138**	-.107**	-.316**	-.241**	-.309**	-.315**	-.148**	-.187**	-.175**	-.243**	-.241**	-.222**
Perceived Effort	PE1	-0.053	0.014	-0.036	0.017	-.108**	-0.015	-.090**	-0.025	-.110**	-.108**	-.125**	-.152**	-.114**	-.130**	-.115**	-.161**	-.168**	-.166**
	PE2	0.03	0.018	0.018	.075**	-0.008	0.021	0.015	0.039	-0.009	-0.02	-0.019	-0.041	-0.025	-0.036	-0.034	-.075**	-.061*	-.082**
	PE3	-.139**	-.060*	-.126**	-0.041	-.165**	-.097**	-.150**	-.095**	-.195**	-.185**	-.160**	-.213**	-.136**	-.151**	-.146**	-.230**	-.239**	-.219**
	PE4	-0.031	-0.006	-0.01	0.03	-.080**	-0.045	-.064*	-0.032	-.077**	-.096**	-.093**	-.095**	-.089**	-.084**	-.068*	-.143**	-.131**	-.141**
Scholarly Altruism	SA1	.335**	.149**	.315**	.196**	.321**	.221**	.313**	.215**	.506**	.391**	.433**	.465**	.237**	.283**	.263**	.354**	.371**	.352**
	SA2	.327**	.135**	.304**	.184**	.306**	.199**	.294**	.206**	.475**	.370**	.401**	.430**	.194**	.241**	.224**	.322**	.346**	.338**
	SA3	.326**	.144**	.304**	.197**	.292**	.183**	.287**	.205**	.474**	.373**	.412**	.446**	.194**	.246**	.215**	.317**	.348**	.332**
	SA4	.314**	.134**	.301**	.150**	.270**	.171**	.246**	.188**	.420**	.352**	.335**	.354**	.195**	.247**	.215**	.274**	.286**	.274**
	SA5	.309**	.137**	.287**	.167**	.252**	.167**	.241**	.185**	.425**	.350**	.343**	.343**	.172**	.209**	.188**	.270**	.283**	.262**
	SA6	.282**	.118**	.270**	.157**	.247**	.156**	.229**	.185**	.417**	.328**	.330**	.323**	.156**	.195**	.178**	.257**	.258**	.244**

Inter-Item and Intra-Item Correlation Matrix (Continued)

		Perceived Career Benefit				Perceived Career Risk				Perceived Effort				Scholarly Altruism					
		PCB1	PCB2	PCB3	PCB4	PCR1	PCR2	PCR3	PCR4	PE1	PE2	PE3	PE4	SA1	SA2	SA3	SA4	SA5	SA6
Perceived Career Benefit	PCB1	**1**																	
	PCB2	**.731****	**1**																
	PCB3	**.653****	**.853****	**1**															
	PCB4	**.687****	**.793****	**.809****	**1**														
Perceived Career Risk	PCR1	-.193**	-.216**	-.210**	-.262**	**1**													
	PCR2	-.154**	-.149**	-.143**	-.205**	**.769****	**1**												
	PCR3	-.188**	-.193**	-.181**	-.216**	**.471****	**.536****	**1**											
	PCR4	-.270**	-.305**	-.300**	-.349**	**.649****	**.697****	**.590****	**1**										
Perceived Effort	PE1	-.124**	-.155**	-.150**	-.180**	.277**	.274**	.354**	.356**	**1**									
	PE2	-0.016	-0.039	-0.033	-0.054	.139**	.130**	.257**	.178**	**.610****	**1**								
	PE3	-.167**	-.206**	-.210**	-.239**	.312**	.292**	.374**	.433**	**.692****	**.548****	**1**							
	PE4	-.102**	-.136**	-.107**	-.148**	.229**	.206**	.321**	.285**	**.652****	**.704****	**.623****	**1**						
Scholarly Altruism	SA1	.306**	.402**	.422**	.419**	-.331**	-.309**	-.291**	-.490**	-.235**	-.055*	-.346**	-.171**	**1**					
	SA2	.302**	.400**	.428**	.426**	-.337**	-.314**	-.284**	-.481**	-.213**	-0.034	-.323**	-.146**	**.908****	**1**				
	SA3	.312**	.396**	.431**	.436**	-.362**	-.342**	-.287**	-.495**	-.212**	-0.029	-.312**	-.154**	**.862****	**.897****	**1**			
	SA4	.292**	.382**	.410**	.405**	-.299**	-.280**	-.220**	-.410**	-.195**	-0.037	-.317**	-.152**	**.652****	**.661****	**.673****	**1**		
	SA5	.306**	.386**	.416**	.408**	-.279**	-.262**	-.208**	-.413**	-.205**	-0.029	-.310**	-.133**	**.671****	**.690****	**.682****	**.853****	**1**	
	SA6	.301**	.382**	.409**	.404**	-.270**	-.257**	-.197**	-.407**	-.196**	-0.036	-.296**	-.127**	**.656****	**.685****	**.667****	**.821****	**.956****	**1**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Appendix 13. $r_{wg(j)}$ for Each Discipline-Level Construct by Discipline

Group-Level Variable Discipline	Regulative Pressure by Funding Agencies	Regulative Pressure by Journals	Normative Pressure	Metadata	Repository
Agricultural Sciences	0.68	0.57	0.64	0.49	0.59
Animal Sciences	0.50	0.57	0.62	0.48	0.57
Anthropology	0.59	0.73	0.62	0.50	0.83
Astronomy	0.69	0.79	0.95	0.75	0.95
Atmospheric Sciences	0.71	0.90	0.94	0.60	0.71
Biochemistry	0.64	0.49	0.86	0.59	0.90
Biology	0.74	0.68	0.85	0.51	0.74
Biomedical Engineering	0.88	0.74	0.71	0.54	0.64
Biometry and Epidemiology	0.80	0.58	0.78	0.61	0.51
Biophysics	0.86	0.31	0.80	0.68	0.87
Botany	0.57	0.45	0.46	0.19	0.37
Cell Biology	0.79	0.74	0.87	0.67	0.87
Chemical Engineering	0.57	0.50	0.63	0.37	0.67
Chemistry	0.69	0.64	0.88	0.54	0.72
Civil Engineering	0.79	0.77	0.80	0.64	0.71
Clinical Psychology	0.63	0.67	0.75	0.23	0.51
Developmental Biology	0.75	0.70	0.85	0.54	0.82
Ecology	0.77	0.70	0.71	0.57	0.69
Electrical Engineering	0.54	0.47	0.81	0.57	0.57
Entomology and Parasitology	0.29	0.60	0.49	0.57	0.72
Environmental Engineering	0.79	0.55	0.78	0.56	0.70
Forestry	0.54	0.83	0.70	0.61	0.68
Genetics	0.71	0.72	0.81	0.53	0.69
Geography	0.55	0.81	0.67	0.38	0.92
Geosciences	0.77	0.76	0.86	0.52	0.76
Mechanical Engineering	0.56	0.69	0.68	0.48	0.62
Metallurgical and Materials Eng.	0.68	0.53	0.74	0.64	0.74
Microbio., Immunology, & Virology	0.73	0.62	0.71	0.42	0.77
Molecular Biology	0.86	0.77	0.85	0.60	0.88
Natural Resources Conservation	0.69	0.78	0.79	0.68	0.60
Neuroscience	0.65	0.65	0.76	0.52	0.75
Nursing	0.67	0.84	0.76	0.63	0.75
Ocean Sciences	0.79	0.65	0.75	0.65	0.75
Oncology/Cancer Research	0.52	0.23	0.60	0.52	0.84
Physics	0.44	0.60	0.53	0.49	0.39
Physiology	0.65	0.63	0.72	0.45	0.63
Plant Sciences	0.60	0.48	0.79	0.58	0.28
Political Science	0.60	0.64	0.78	0.34	0.60
Preventive Med. & Comm. Health	0.28	0.85	0.65	0.67	0.70
Psychology, Combined	0.44	0.33	0.57	0.32	0.38
Psychology, Except Clinical	0.63	0.60	0.68	0.59	0.64
Public Administration	0.75	0.82	0.77	0.66	0.80
Sociology	0.51	0.87	0.77	0.51	0.67
Average $r_{wg(j)}$	0.65	0.65	0.74	0.53	0.69
Median $r_{wg(j)}$	0.67	0.65	0.76	0.54	0.70
Minimum $r_{wg(j)}$	0.28	0.23	0.46	0.19	0.28
Maximum $r_{wg(j)}$	0.88	0.90	0.95	0.75	0.95

References

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

Ajzen, I. 2002. "Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior," *Journal of Applied Social Psychology* (32:4), pp 665-683.

Ajzen, I., and Fishbein, M. 1980. *Uncerstanding Attitudes and Predicting Social Behavior*, Prentice-Hall: Englewood Cliffs, NJ.

Ajzen, I., Kuhl, J., and Beckmann, J. 1985. *From intentions to actions: A theory of planned behavior*, Springer: New York.

Akbulut-Bailey, A. 2011. "Information sharing between local and state governments," *Journal of Computer Information Systems* (51:4), pp 53-63.

Allison, P. 1999. *Multiple Regession: A Primer*, Pine Forge Press: Thousand Oaks, CA.

Andersen, H. 2001. "The norm of universalism in sciences: Social origin and gender of researchers in Denmark," *Scientometrics* (50:255-272).

Armitage, C. J., and Conner, M. 1999. "Distinguishing perceptions of control from self-efficacy: Predicting consumption of a low-fat diet using the theory of planned behavior," *Journal of Applied Social Psychology* (29:1), pp 72-90.

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

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P., and Wright, M. H. 2003. "Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure."

Awad, N. F., and Krishnan, M. S. 2006. "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization," *Mis Quarterly* (30:1) Mar, pp 13-28.

Babbie, E. R. 1990. *Survey Research Methods*, Wadsworth Publishing: Belmont, CA.

Baker, K. S., and Bowker, G. C. 2007. "Information ecology: open system environment for data, memories, and knowing," *Journal of Intelligent Information Systems* (29:1) Aug, pp 127-144.

Baker, K. S., and Yarmey, L. 2009. "Data Stewardship: Environmental Data Curation and a Web-of-Repositories," *International Journal of Digital Curation* (2:4), pp 12-27.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*, Prentice-Hall: Englewood Cliffs, NJ.

Barber, B. 1952. *Science and the social order*, Free Press: Glencoe, IL.

Barley, S. R. 1986. "Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments," *Administrative Science Quarterly* (31:1) Mar, pp 78-108.

Barley, S. R., and Tolbert, P. S. 1997. "Institutionalization and structuration: Studying the links between action and institution," *Organization Studies* (18:1) 1997, pp 93-117.

Batson, C. D. 1991. *The altruism question: Toward a social-psychological answer*, Lawrence Erlbaum Associates, Inc: Hillsdale, NJ.

Battilana, J. 2006. "Agency and institutions: The enabling role of individuals' social position," *Organization* (13:5) Sep, pp 653-676.

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

Bebeau, M. J., and Monson, V. 2011. "Authorship and Publication Practices in the Social Sciences: Historical Reflections on Current Practices," *Science and Engineering Ethics* (17:2) Jun, pp 365-388.

Benbasat, I., Goldstein, D., and Mead, M. 1987. "The case research strategy in studies of information systems," *MIS Quarterly* (11:3), pp 369-386.

Benlian, A., and Hess, T. 2011. "Opportunities and risks of software-as-a-service: Findings from a survey of IT executives," *Decision Support Systems* (52:1) Dec, pp 232-246.

Berry, W. D., and Feldman, S. 1985. *Multiple Regression in Practice*, SAGE Publications: Thousand Oaks.

Bickel, R. 2007. *Multilevel Analysis for Applied Research*, Guilford Press: New York, NY.

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

Birnholtz, J. P., and Bietz, M. J. Year. "Data at work: Supporting sharing in science and engineering," 2003, pp. 339-348.

Bjorck, F. Year. "Institutional theory: A new perspective for research into IS/IT security in organisations," Proceedings of the 37th Hawaii International Conference on System Sciences, Citeseer2004, p. 190.

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 and S. W. l. Kozlowski (eds.), Jossey Bass, Inc.: San Francisco, CA.

Bloor, D. 1976. *Knowledge and social imagery*, Routledge & K. Paul: London; Boston.

Blumenthal, D., Campbell, E. G., Anderson, M. S., Causino, N., and Louis, K. S. 1997. "Withholding research results in academic life science - Evidence from a national survey of faculty," *Jama-Journal of the American Medical Association* (277:15) Apr, pp 1224-1228.

Blumenthal, D., Campbell, E. G., Causino, N., and Louis, K. S. 1996. "Participation of life-science faculty in research relationships with industry," *New England Journal of Medicine* (335:23) Dec, pp 1734-1739.

Blumenthal, D., Campbell, E. G., Gokhale, M., Yucel, R., Clarridge, B., Hilgartner, S., and Holtzman, N. A. 2006. "Data withholding in genetics and the other life sciences: Prevalences and predictors," *Academic Medicine* (81:2) Feb, pp 137-145.

Bock, G. W., Zmud, R. W., Kim, Y. G., and Lee, J. N. 2005. "Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate," *MIS Quarterly: Management Information Systems* (29:1), pp 87-111.

Borgman, C. L. 1990. "Editor's introduction," in *Scholarly communication and bibliometrics,* C. L. Borgman (ed.), Sage Publications: Newbury Park, CA, pp. 10-27.

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

Borgman, C. L. 2009. "The digital future is now: A call to action for the humanities," *Digital Humanities Quarterly* (3:4).

Borgman, C. L. 2010. "Research Data: Who will share what, with whom, when, and why?," in *Fifth China - North America Library Conference*: Beijing, China, pp. 1-21.

Borgman, C. L., Wallis, J. C., and Enyedy, N. 2007a. "Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries," *International Journal on Digital Libraries* (7:1-2), pp 17-30.

Borgman, C. L., Wallis, J. C., Mayernik, M. S., and Pepe, A. Year. "Drowning in data: Digital library architecture to support scientific use of embedded sensor networks," 2007b, pp. 269-277.

Bosnjak, M., Tuten, T. L., and Wittmann, W. W. 2005. "Unit (non) response in web-based access panel surveys: An extended planned-behavior approach," *Psychology & Marketing* (22:6), pp 489-505.

Bourne, P. 2005. "Will a biological database be different from a biological journal?," *Plos Computational Biology* (1:3) Aug, pp 179-181.

Bowker, G., and Star, S. L. 2000. *Sorting Things Out*, MIT Press: Cambridge, MA.

Bowker, G. C., and Star, S. L. 1999. *Sorting Things Out: Classification and Its Consequences*, MIT Press: Cambridge, MA.

Boyce, D., Judson, B., and Hall, S. 2006. "Data sharing - A case of shared databases and community use of on-line GIS support systems," *Environmental Monitoring and Assessment* (113:385-394).

Brandt, D. S. 2007. "Librarians as partners in e-research: Purdue University Libraries promote collaboration," *College and Research Libraries News* (68:6), pp 365-367+396.

Braxton, J. M. 1986. "The normative structure of science: Social control in the academic profession," in *Higher education: Handbook of theory and research,* J. C. Smart (ed.), Agathon Press: New York, NY, pp. 309-357.

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) Aug, pp 926-938.

Budros, A. 2002. "The mean and lean firm and downsizing: Causes of involuntary and voluntary downsizing strategies," *Sociological Forum* (17:2) Jun, pp 307-342.

Buetow, K. H. 2005. "Cyberinfrastructure: Empowering a "third way" in biomedical research," *Science* (308:5723) May, pp 821-824.

Burkell, J. 2003. "The Dilemma of Survey Nonresponse," *Library & Information Science Research* (25:3), pp 239-263.

Burt, R. S. 1982. *Toward a Structural Theory of Action: Network Models of Social Structure, Perception, and Action*, Academic Press: New York.

Burt, R. S. 1987. "Social Contagion and Innovation - Cohesion versus Structural Equivalence," *American Journal of Sociology* (92:6) May, pp 1287-1335.

Burton-Jones, A., and Straub, D. W. 2006. "Reconceptualizing System Usage: An Approach and Empirical Test," *Information Systems Research* (17:3), pp 228-246.

Campbell, E. G., and 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), pp 241-255.

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

Campbell, E. G., Louis, K. S., and 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) Apr, pp 995-999.

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

Carlson, S. 2006. "Lost in a sea of science data," in *The Chronicle of Higher Education*.

Carlson, S., and 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) Jan.

Carney, M., Gedajlovic, E., and Yang, X. H. 2009. "Varieties of Asian capitalism: Toward an institutional theory of Asian enterprise," *Asia Pacific Journal of Management* (26:3) Sep, pp 361-380.

Cech, T. R., Eddy, S. R., Eisenberg, D., Hersey, K., Holtzman, S. H., Poste, G. H., Raikhel, N. V., Scheller, R. H., Singer, D. B., Waltham, M. C., and Comm Responsibilities Authorship, B. 2003. "Sharing publication-related data and materials: Responsibilities of authorship in the life sciences," *Plant Physiology* (132:1) May, pp 19-24.

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

Cheung, W., Chang, M. K., and Lai, V. S. 2000. "Prediction of Internet and World Wide Web usage at work: a test of an extended Triandis model," *Decision Support Systems* (30:1), pp 83-100.

Chiasson, M. W., and Davidson, E. 2005. "Taking industry seriously in information systems research," *MIS Quarterly* (29:4) Dec, pp 591-605.

Chiu, C.-M., Hsu, M.-H., and Wang, E. T. 2006. "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories," *Decision support systems* (42:3), pp 1872-1888.

Cho, H., Chen, M., and Chung, S. 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) Jun, pp 1198-1212.

Cho, V. 2006. "A study of the roles of trusts and risks in information-oriented online legal services using an integrated model," *Information & Management* (43:4), pp 502-520.

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

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates: Mahwah, NJ.

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

Colditz, G. A. 2009. "Constraints on Data Sharing Experience From the Nurses' Health Study," *Epidemiology* (20:2) Mar, pp 169-171.

Cole, J., and Cole, S. 1973. *Social Stratification in Science*, University of Chicago Press: Chicago.

Collins, H. M. 1992. *Changing order: Replication and induction in scientific practice*, University of Chicago Press: Chicago, IL.

Conchar, M. P., Zinkhan, G. M., Peters, C., and Olavarrieta, S. 2004. "An integrated framework for the conceptualization of consumers' perceived-risk processing," *Journal of the Academy of Marketing Science* (32:4) Fal, pp 418-436.

Constant, D., Kiesler, S., and Sproull, L. 1994. "What's mine is ours, or is it - A study of attitudes about information sharing," *Information Systems Research* (5:4) Dec, pp 400-421.

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

Contento, I. R. 2011. *Nutrition Education: Linking Research, Theory, and Practice*, Jones and Bartlett Publishers: Sudbury, MA.

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

Cragin, M. H., and Shankar, K. 2006. "Scientific data collections and distributed collective practice," *Computer Supported Cooperative Work: CSCW: An International Journal* (15:2-3), pp 185-204.

Craig, J. R., and Reese, S. C. 1973. "Psychology in action - Retention of raw data - problem revisited," *American Psychologist* (28:8), pp 723-723.

Crall, A. W., Newman, G. J., Jarnevich, C. S., Stohlgren, T. J., Waller, D. M., and Graham, J. 2010. "Improving and integrating data on invasive species collected by citizen scientists," *Biological Invasions* (12:10) Oct, pp 3419-3428.

Creswell, J. W. 2008. *Research Design: Qualitative, Quantitative, and Mixed methods Approaches*, SAGE Publications: Thousand Oaks.

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

Daniels, K., Johnson, G., and de Chernatony, L. 2002. "Task and institutional influences on managers' mental models of competition," *Organization Studies* (23:1) 2002, pp 31-62.

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

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

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance in Information Technology," *MIS Quarterly* (13:3), pp 319-340.

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1989. "User acceptance of computer technology: a comparison of two theoretical models," *Management Science* (35:8), pp 982-1003.

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

Deephouse, D. L. 1996. "Does isomorphism legitimate?," *Academy of Management Journal* (39:4) Aug, pp 1024-1039.

Delserone, L. M. 2008. "At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries," *Library Trends* (57:2) Fal, pp 202-210.

DeVellis, R. F. 2003. *Scale Development: Theory and Applications*, Sage Publications: Thousand Oaks: CA.

Diamond, A. M. 1986. "What is a citation worth," *Journal of Human Resources* (21:2) Spr, pp 200-215.

Diaz, L., Granell, C., Gould, M., and Huerta, J. 2011. "Managing user-generated information in geospatial cyberinfrastructures," *Future Generation Computer Systems-the International Journal of Grid Computing-Theory Methods and Applications* (27:3) Mar, pp 304-314.

Dickinger, A., Arami, M., and Meyer, D. 2008. "The role of perceived enjoyment and social norm in the adoption of technology with network externalities," *European Journal of Information Systems* (17:1), pp 4-11.

Dillman, D. A. 2007. *Mail and Internet Surveys: The Tailored Design Method (2nd Ed.)*, John & Sons, Inc.: Hoboken, NJ.

DiMaggio, P. J., and Powell, W. W. 1983. "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields," *American Sociological Review* (48:2) 1983, pp 147-160.

DiMaggio, P. J., and Powell, W. W. 1991. "Introduction," in *The New Institutionalism in Organizational Analysis,* W. W. Powell and P. J. DiMaggio (eds.), The University of Chicago Press.: Chicago: The University of Chicago Press., pp. 1-38.

Dixon, M. A., and Cunningham, G. B. 2006. "Data aggregation in multilevel analysis: a review of conceptual and statistical issues," *Measurement in Physical Education and Exercise Science* (10:2), pp 85-107.

Doll, W. J., and Torkzadeh, G. 1988. "The measurement of end-user computing satisfaction," *MIS quarterly* (12:2), pp 259-274.

Dundar, H., and Lewis, D. R. 1998. "Determinants of research productivity in higher education," *Research in Higher Education* (39:6), pp 607-631.

Duxbury, L., and Haines, G. 1991. "Predicting alternative work arrangements from salient attitudes - A study of decision makers in the public-sector," *Journal of Business Research* (23:1) Aug, pp 83-97.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., and Borgman, C. L. 2011. "Science friction: Data, metadata, and collaboration," *Social Studies of Science* (41:5) Oct, pp 667-690.

Eschenfelder, K., and Johnson, A. 2011. "The limits of sharing: Controlled data collections," in *Annual Meeting of the American Society for Information Science and Technology*: New Orleans, Louisiana

Ethington, C. A. 1997. "A hierarchical linear modeling approach to studying college effects," in *Higher education: Handbook of theory and research,* J. C. Smart (ed.), Agathon Press: New York, NY, pp. 165-194.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. 1999. "Evaluating the use of exploratory factor analysis in psychological research," *Psychological methods* (4:3), pp 272-299.

Faniel, I. M. 2009. "Unrealized potential: The socio-technical challenges of a large scale cyberinfrastructure initiative."

Faniel, I. M., and Jacobsen, T. E. 2010. "Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data," *Computer Supported Cooperative Work-the Journal of Collaborative Computing* (19:3-4) Aug, pp 355-375.

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

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

Fehr, E., and Fischbacher, U. 2003. "The nature of human altruism," *Nature* (425:6960) Oct 23, pp 785-791.

Fehr, E., and Schmidt, K. M. 2006. "The economics of fairness, reciprocity and altruism–experimental evidence and new theories," in *Handbook of the Economics of Giving, Altruism and Reciprocity,* S.-C. Kolm and J. M. Ythier (eds.), pp. 615-691.

Fennema-Notestine, C. 2009. "Enabling public data sharing: encouraging scientific discovery and education," *Methods in molecular biology (Clifton, N.J.)* (569), pp 25-32.

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

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M. J., Angiuoli, S. V., Ashburner, M., Axelrod, N., Baldauf, S., Ballard, S., Boore, J., Cochrane, G., Cole, J., Dawyndt, P., De Vos, P., Depamphilis, C., Edwards, R., Faruque, N., Feldman, R., Gilbert, J., Gilna, P., Glöckner, F. O., Goldstein, P., Guralnick, R., Haft, D., Hancock, D., Hermjakob, H., Hertz-Fowler, C., Hugenholtz, P., Joint, I., Kagan, L., Kane, M., Kennedy, J., Kowalchuk, G., Kottmann, R., Kolker, E., Kravitz, S., Kyrpides, N., Leebens-Mack, J., Lewis, S. E., Li, K., Lister, A. L., Lord, P., Maltsev, N., Markowitz, V., Martiny, J., Methe, B., Mizrachi, I., Moxon, R., Nelson, K., Parkhill, J., Proctor, L., White, O., Sansone, S. A., Spiers, A., Stevens, R., Swift, P., Taylor, C., Tateno, Y., Tett, A., Turner, S., Ussery, D., Vaughan, B., Ward, N., Whetzel, T., San Gil, I., Wilson, G., and Wipat, A. 2008. "The minimum information about a genome sequence (MIGS) specification," *Nature Biotechnology* (26:5), pp 541-547.

Fienberg, S. E. 1994. "Sharing statistical data in the biomedical and health sciences: Ethical, institutional, legal, and professional dimensions," *Annual Review of Public Health* (15), pp 1-18.

Fienberg, S. E., Martin, M. E., and Straf, M. L. 1985. *Sharing Research Data*, National Academy Press: Washington, D.C.

Fishbein, M. 1980. "A theory of reasoned action: some applications and implications," *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation* (27), pp 65-116.

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

Fisher, J. B., and Fortmann, L. 2010. "Governing the data commons: Policy, practice, and the advancement of science," *Information & Management* (47:4) May, pp 237-245.

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

Fox, J. 1991. *Regression diagnostics*, Sage Publications: Newbury Park, CA.

Friedland, R., and Alford, R. R. 1991. "Bringing society back in: Practices, and institutional contradictions," in *The New Institutionalism in Organizational Analysis,* W. W. Powell and P. J. DiMaggio (eds.), University of Chicago Press: Chicago, pp. 232-263.

Garud, R., Jain, S., and Kumaraswamy, A. 2002. "Institutional entrepreneurship in the sponsorship of common technological standards: The case of Sun Microsystems and Java," *Academy of Management Journal* (45:1) Feb, pp 196-214.

Garvey, W. D. 1979. *Communication: The essence of science: Facilitating the exchange among librarians, scientists, engineers, and students*, Pergamon Press: Oxford.

Gaston, J. 1973. *Originality and Competition in Science: A Study of the British High Energy Physics Community*, University of Chicago Press: Chicago.

Gauch, H. G. 2003. *Scientific method in practice*, Cambridge University Press: Cambridge, UK.

Gefen, D., Straub, D. W., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the AIS* (4:7), pp 1-77.

George, E., Chattopadhyay, P., Sitkin, S. B., and Barden, J. Q. 2006. "Cognitive underpinnings of institutional persistence and change: A framing perspective," *Academy of Management Review* (31:2) Apr, pp 347-365.

Giffels, J. 2010. "Sharing Data is a Shared Responsibility: Commentary on: "The Essential Nature of Sharing in Science"," *Science and Engineering Ethics* (16:4), pp 801-803.

Glover, D. M., Chandler, C. L., Doney, S. C., Buesseler, K. O., Heimerdinger, G., Bishop, J. K. B., and Flierl, G. R. 2006. "The US JGOFS data management experience," *Deep-Sea Research Part Ii-Topical Studies in Oceanography* (53:5-7), pp 793-802.

Goldsmith, M. 1967. "The autonomy of science: Some thoughts for discussion," *The Political Quarterly* (38:1), pp 81-89.

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

Granfield, R. 2007. "The meaning of pro bono: Institutional variations in professional obligations among lawyers," *Law & Society Review* (41:1) Mar, pp 113-146.

Greene, J. C., Caracelli, V. J., and Graham, W. F. 1989. "Toward a conceptual framework for mixed-method evaluation designs," *Educational Evaluation and Policy Analysis* (11:3), pp 255-274.

Greenwood, R., and Suddaby, R. 2006. "Institutional entrepreneurship in mature fields: The big five accounting firms," *Academy of Management Journal* (49:1) Feb, pp 27-48.

Grewal, R., and Dharwadkar, R. 2002. "The role of the institutional environment in marketing channels," *Journal of Marketing* (66:3) Jul, pp 82-97.

Groves, R. M. 2006. "Nonresponse rates and nonresponse bias in household surveys," *Public Opinion Quarterly* (70:5), pp 646-675.

Groves, R. M., and Peytcheva, E. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias A Meta-Analysis," *Public Opinion Quarterly* (72:2), pp 167-189.

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

Hall, P. A., and Taylor, R. C. R. 1996. "Political science and the three new institutionalisms," *Political Studies* (44:5) Dec, pp 936-957.

Haveman, H. A. 1993. "Follow the leader: mimetic isomorphism and entry into new markets," *Administrative Science Quarterly* (38:4) Dec, pp 593-627.

Haveman, H. A., and Rao, H. 1997. "Structuring a theory of moral sentiments: Institutional and organizational coevolution in the early thrift industry," *American Journal of Sociology* (102:6) May, pp 1606-1651.

He, W., and Wei, K.-K. 2009. "What drives continued knowledge sharing? An investigation of knowledge-contribution and -seeking beliefs," *Decision Support Systems* (46:4) Mar, pp 826-838.

Heck, R., and Thomas, S. 1999. *An Introduction to Multilevel Modeling Techniques*, Lawrence Erlbaum Associates: Mahwah, NJ.

Heidorn, P. B. 2008. "Shedding light on the dark data in the long tail of science," *Library Trends* (57:2), pp 280-299.

Heinrich, C. J., and Fournier, E. 2004. "Dimensions of publicness and performance in substance abuse treatment organizations," *Journal of Policy Analysis and Management* (23:1) Win, pp 49-70.

Heugens, P. P. M. A. R., and Lander, M. W. 2009. "Structure! Agency! (and other quarrels): a meta-analysis of institutional theories of organization," *Academy of Management Journal* (52:1) Feb, pp 61-85.

Hey, T., and Hey, J. 2006. "e-Science and its implications for the library community," *Library Hi Tech* (24:4), pp 515-528.

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

Hey, T., and Trefethen, A. E. 2002. "The UK e-science core programme and the grid," *Future Generation Computer Systems* (18:8) Oct, pp 1017-1031.

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

Hey, T., Trefethen, A. 2003. "The data deluge: An e-science perspective," in *Grid computing: Making the global infrastructure a reality,* F. Berman, G. C. Fox and T. Hey (eds.), Wiley: New York, pp. 809-824.

Hilgartner, S., and Brandtrauf, S. I. 1994. "Data access, ownership, and control - Toward empirical-studies of access practices," *Knowledge-Creation Diffusion Utilization* (15:4) Jun, pp 355-372.

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

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., and Zaslavsky, I. 2011. "Components of an environmental observatory information system," *Computers & Geosciences* (37:2) Feb, pp 207-218.

Howell, D. C. 2012. *Statistical Methods for Psychology*, Wadsworth Publishing Company: Belmont, CA.

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

Hsu, C.-L., and Lin, J. C.-C. 2008. "Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation," *Information & Management* (45:1) Jan, pp 65-74.

Hsu, M.-H., and Chiu, C.-M. 2004. "Predicting electronic service continuance with a decomposed theory of planned behaviour," *Behaviour & Information Technology* (23:5), pp 359 - 373.

Hu, P. J.-H., Chau, P. Y., and Sheng, O. R. L. 2002. "Adoption of telemedicine technology by health care organizations: an exploratory study," *Journal of organizational computing and electronic commerce* (12:3), pp 197-221.

Hung, S.-Y., Durcikova, A., Lai, H.-M., and Lin, W.-M. 2011a. "The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior," *International Journal of Human-Computer Studies* (69:6) Jun, pp 415-427.

Hung, S.-Y., Lai, H.-M., and Chang, W.-W. 2011b. "Knowledge-sharing motivations affecting R&D employees' acceptance of electronic knowledge repository," *Behaviour & Information Technology* (30:2), pp 213-230.

Husted, K., and Michailova, S. 2002. "Diagnosing and fighting knowledge-sharing hostility," *Organizational Dynamics* (31:1), pp 60-73.

Hwang, Y., Erkens, D. H., and Evans III, J. H. 2009. "Knowledge sharing and incentive design in production environments: Theory and evidence," *The Accounting Review* (84:4), pp 1145-1170.

Inkpen, A. C. 1996. "Creating knowledge through collaboration," *California Management Review* (39:1) Fal, pp 123-&.

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

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

Jasperson, J., Carter, P. E., and Zmud, R. W. 2005. "A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated With Information Technology Enabled Work Systems," *MIS Quarterly* (29:3), pp 525-557.

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C., and Voss, A. 2005. "Collaboration and trust in healthcare innovation: The eDiaMoND case study," *Computer Supported Cooperative Work: CSCW: An International Journal* (14:4), pp 369-398.

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

Johnson, T., Adams, D., and Kim, J. S. 2010. "Mapping the perspectives of low-income parents in a children's college savings account program," *Children and Youth Services Review* (32:1) Jan, pp 129-136.

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

Kankanhalli, A., Tan, B. C. Y., and Wei, K. K. 2005. "Contributing knowledge to electronic knowledge repositories: An empirical investigation," *MIS Quarterly: Management Information Systems* (29:1), pp 113-143.

Karasti, H., and Baker, K. S. 2008. "Digital Data Practices and the Long Term Ecological Research Program Growing Global," *International Journal of Digital Curation* (2:3), pp 42-58.

Karasti, H., Baker, K. S., and Halkola, E. 2006. "Enriching the notion of data curation in e-Science: Data managing and information infrastructuring in the Long Term Ecological Research (LTER) network," *Computer Supported Cooperative Work: CSCW: An International Journal* (15:4), pp 321-358.

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

Ke, W., Liu, H., Wei, K. K., Gu, J., and Chen, H. 2009. "How do mediated and non-mediated power affect electronic supply chain management system adoption? The mediating effects of trust and institutional pressures," *Decision Support Systems* (46:4), pp 839-851.

Khatibi, V., and Montazer, G. A. 2009. "E-Research Process Framework," *World Academy of Science, Engineering and Technology* (54), pp 386-391.

Kim, B., and Han, I. 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), pp 1012-1026.

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

Kim, S. S., and Malhotra, N. K. 2005. "A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Postadoption Phenomena," *Management Science* (51:5), pp 741-755.

Kim, Y. S., and Stanton, J. M. 2012. "Institutional and Individual Influences on Scientists' Data Sharing Practices," *Journal of Computational Science Education* (3:1), pp 47-56.

Kisfalvi, V., and Maguire, S. 2011. "On the Nature of Institutional Entrepreneurs: Insights From the Life of Rachel Carson," *Journal of Management Inquiry* (20:2) Jun, pp 152-177.

Klein, K. J., Dansereau, F., and Hall, R. J. 1994. "Levels issues in theory development, data collection, and analysis," *Academy of Management Review* (19), pp 195-229.

Klein, K. J., and Kozlowski, S. W. J. 2000. "From Micro to Meso: Critical Steps in Conceptualizing and Conducting Multilevel Research," *Organizational Research Methods* (3:3), pp 211-236.

Klein, R. 2007. "An empirical examination of patient-physician portal acceptance," *European Journal of Information Systems* (16:6), pp 751-760.

Kline, R. B. 2005. *Principles and Practice of Structural Equation Modeling (2nd ed.)*, The Guilford Press: New York, NY.

Kling, R., McKim, G., Fortuna, J., and King, A. 2000. "Scientific Collaboratories as Socio-Technical Interaction Networks: A Theoretical Approach. School of Library and Information Science."

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

Kolekofski Jr, K. E., and Heminger, A. R. 2003. "Beliefs and attitudes affecting intentions to share information in an organizational setting," *Information and Management* (40:6), pp 521-532.

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

Koulikoff-Souviron, M., and Harrison, A. 2008. "Interdependent supply relationships as institutions: The role of HR practices," *International Journal of Operations and Production Management* (28:5), pp 412-432.

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

Krathwohl, D. R. 1998. *Methods of educational and social science research: An integrated approach*, Longman: New York.

Krebs, D. 1975. "Empathy and altruism," *Journal of Personality and Social Psychology* (32:6) 1975, pp 1134-1146.

Kreft, I., and De Leeuw, J. 1998. *Introducing Multilevel Modeling*, Sage Publications: Thousand Oaks.

Kuhn, T. S. 1996. *The Structure of Scientific Revolutions (3rd Ed)*, University of Chicago Press: Chicago.

Kuo, F. Y., and Young, M. L. 2008a. "Predicting knowledge sharing practices through intention: A test of competing models," *Computers in Human Behavior* (24:6) Sep, pp 2697-2722.

Kuo, F. Y., and Young, M. L. 2008b. "A study of the intention - Action gap in knowledge sharing practices," *Journal of the American Society for Information Science and Technology* (59:8) Jun, pp 1224-1237.

Lam, A. 2010. "From 'Ivory Tower Traditionalists' to 'Entrepreneurial Scientists'? Academic Scientists in Fuzzy University-Industry Boundaries," *Social Studies of Science* (40:2) Apr, pp 307-340.

Lane, J., and Schur, C. 2010. "Balancing Access to Health Data and Privacy: A Review of the Issues and Approaches for the Future," *Health Services Research* (45:5), pp 1456-1467.

Latour, B. 1987. *Science in action: How to follow scientists and engineers though society*, Harvard University Press: Cambridge, MA.

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

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

Lee, G., and Lee, W. J. 2010. "Altruistic traits and organizational conditions in helping online," *Computers in Human Behavior* (26:6), pp 1574-1580.

Lee, J., and Rao, H. R. 2009. "Task complexity and different decision criteria for online service acceptance: A comparison of two e-government compliance service domains," *Decision Support Systems* (47:4) Nov, pp 424-435.

Levy, Y. 2006. *Assessing the value of e-learning systems*, Information Science Publishing: Hershey, PA.

Liang, H. G., Saraf, N., Hu, Q., and Xue, Y. J. 2007. "Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management," *MIS Quarterly* (31:1) Mar, pp 59-87.

Limayem, M., Hirt, S. G., and Cheung, C. M. K. 2007. "How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance," *MIS Quarterly* (31:4), pp 705-737.

Lin, C.-P. 2008. "Clarifying the relationship between organizational citizenship behaviors, gender, and knowledge sharing in workplace organizations in Taiwan," *Journal of Business and Psychology* (22:3) Mar, pp 241-250.

Lin, H.-F. 2007. "Effects of extrinsic and intrinsic motivation on employee knowledge sharing intentions," *Journal of Information Science* (33:2) 2007, pp 135-149.

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

Liotta, L. A., Lowenthal, M., Mehta, A., Conrads, T. P., Veenstra, T. D., Fishman, D. A., and Petricoin Iii, E. F. 2005. "Importance of communication between producers and consumers of publicly available experimental data," *Journal of the National Cancer Institute* (97:4), pp 310-314.

Liu, C., Sia, C.-L., and Wei, K.-K. 2008. "Adopting organizational virtualization in B2B firms: An empirical study in Singapore," *Information & Management* (45:7), pp 429-437.

Liu, H., Ke, W., Wei, K. K., Gu, J., and 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), pp 372-384.

Louis, K. S., Jones, L. M., and Campbell, E. G. 2002. "Sharing in science," *American Scientist* (90:4) Jul-Aug, pp 304-307.

Luo, X. W. 2007. "Continuous learning: The influence of national institutional logics on training attitudes," *Organization Science* (18:2) Mar-Apr, pp 280-296.

Maas, C. J., and Hox, J. J. 2005. "Sufficient sample sizes for multilevel modeling," *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* (1:3), pp 86-92.

MacFarlane, B., and Cheng, M. 2008. "Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms," *Journal of Academic Ethics* (6:1), pp 67-78.

Manstead, A. S. R., and Van Eekelen, S. A. M. 1998. "Distinguishing between perceived behavioral control and self-efficacy in the domain of academic achievement intentions and behaviors," *Journal of Applied Social Psychology* (28:15), pp 1375-1392.

Marcial, L. H., and 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) Oct, pp 2029-2048.

Markus, M. L. 2001. "Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success," *Journal of Management Information Systems* (18:1) Sum, pp 57-93.

Marshall, E. 1990. "Data sharing: A declining ethic?," *Science* (248:4958), pp 952-957.

Mayer, D., Nishii, L., Schneider, B., and Goldstein, H. 2007. "The precursors and products of justice climates: Group leader antecedents and employee attitudinal consequences," *Personnel Psychology* (60:4), pp 929-963.

McCain, K. W. 1991. "Communication, competition, and secrecy - The production and dissemination of research-related information in genetics," *Science Technology & Human Values* (16:4) Fal, pp 491-516.

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

McCain, K. W. 2000. "Sharing digitized research-related information on the World Wide Web," *Journal of the American Society for Information Science* (51:14) Dec, pp 1321-1327.

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

McGrath, P. J. 2002. *Scientists, business, and the state, 1890-1960*, (University of North Carolina Press: Chapel Hill.

McLure Wasko, M., and 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), pp 155-173.

Merton, R. K. 1957. "Priorities in scientific discovery," in *The Sociology of Science,* N. W. Storer (ed.), University of Chicago Press: Chicago, pp. 286-324.

Merton, R. K. 1968. *Social theory and social structure*, Free Press: New York.

Merton, R. K. 1970. *Science, technology & society in seventeenth century England (1st ed.)*, (H. Fertig: New York.

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

Merton, R. K., and Sztompka, P. 1996. *On social structure and science*, University of Chicago Press: Chicago, IL.

Meyer, J. W., and Rowan, B. 1977. "Institutionalized organizations: Formal structure as myth and ceremony," *American Journal of Sociology* (83:2) 1977, pp 340-363.

Mezias, S. J., and Scarselletta, M. 1994. "Resolving financial-reporting problems - An institutional analysis of the process," *Administrative Science Quarterly* (39:4) Dec, pp 654-678.

Millerand, F., and Baker, K. S. 2010. "Who are the users? Who are the developers? Webs of users and developers in the development process of a technical standard," *Information Systems Journal* (20:2) Mar, pp 137-161.

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), pp 569-595.

Miyazaki, A. D., and Fernandez, A. 2001. "Consumer perceptions of privacy and security risks for online shopping," *Journal of Consumer Affairs* (35:1) Sum, pp 27-44.

Moore, G. C., and Benbasat, I. 1991. "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Information Systems Research* (2:3), pp 192-222.

Mowery, D. C. 2005. "The Bayh-Dole act and high-technology entrepreneurship in U.S. universities: Chicken, egg, or something else?," in *University entrepreneurship and technology transfer: Process, design, and intellectual property,* G. Libecap (ed.), Elsevier: Amsterdam, Netherlands, pp. 39-68.

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

Nahapiet, J., and Ghoshal, S. 1998. "Social capital, intellectual capital, and the organizational advantage," *Academy of Management Review* (23:2) Apr, pp 242-266.

National Institutes of Health 2003. "NIH Data Sharing Policy and Implementation Guidance."

National Science Foundation 2008. "Data Archiving Policy."

National Science Foundation 2010. "Scientists seeking NSF funding will soon be required to submit data management plans."

Netemeyer, R. G., Boles, J. S., and McMurrian, R. C. 1996. "Development and validation of work-family conflict and family-work conflict scales," *Journal of Applied Psychology* (81:4), pp 400-410.

Neufeld, D. J., Dong, L., and Higgins, C. 2007. "Charismatic leadership and user acceptance of information technology," *European Journal of Information Systems* (16:4), pp 494-510.

Noor, M. A. F., Zimmerman, K. J., and Teeter, K. C. 2006. "Data sharing: How much doesn't get submitted to GenBank?," *Plos Biology* (4:7) Jul, pp 1113-1114.

Nunnally , J. C., and Bernstein, I. H. 1994. *Psychometric Theory (3rd ed.)*, McGraw-Hill: New York, NY.

Oliver, C. 1991. "Strategic responses to institutional processes," *Academy of Management Review* (16:1) Jan, pp 145-179.

Oliver, C. 1997. "Sustainable competitive advantage: Combining institutional and resource-based views," *Strategic Management Journal* (18:9) Oct, pp 697-713.

Ostroff, C., and Harrison, D. A. 1999. "Meta-analysis, level of analysis, and best estimates of population correlations: Cautions for interpreting meta-analytic results in organizational behavior," *Journal of Applied Psychology* (84:2), pp 260-270.

Paton, N. W. 2008. "Managing and sharing experimental data: standards, tools and pitfalls," *Biochemical Society Transactions* (36) Feb, pp 33-36.

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) Spr, pp 101-134.

Pavlou, P. A., and Fygenson, M. 2006. "Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior," *MIS Quarterly: Management Information Systems* (30:1), pp 115-143.

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

Phang, C. W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B. C. Y., and Teo, H.-H. 2006. "Senior Citizens' Acceptance of Information Systems: A Study in the Context of e-Government Services," *IEEE Transaction on Engineering Management* (53:4), pp 555-569.

Phillips, N., and Tracey, P. 2007. "Opportunity recognition, entrepreneurial capabilities and bricolage: connecting institutional theory and entrepreneurship in strategic organization," *Strategic Organization* (5:3) Aug, pp 313-320.

Pienta, A. M., Alter, G. C., and Lyle, J. A. 2010. "The enduring value of social science research: The use and reuse of primary research data," in *The Organisation, Economics and Policy of Scientific Research*: Torino, Italy.

Pierce, S. J. 1990. "Disciplinary work and interdisciplinary areas: Sociology and bibliometrics," in *Scholarly communication and bibliometrics,* C. L. Borgman (ed.), Sage Publications: Newbury Park, CA, pp. 46-58.

Pinch, T. J., and Bijker, W. E. 1984. "The social construction of facts and artefacts: or How the sociology of science and the sociology of technology might benefit each other," *Social Studies of Science* (14:3), pp 399-441.

Pinelli, T. E. 1991. "The information-seeking habits and practices of engineers," *Science & Technology Libraries* (11:3), pp 5-25.

Piwowar, H. A. 2010. *Foundational studies for measuring the impact, prevalence, and patterns of publicly sharing biomedical research data*, University of Pittsburgh.

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

Piwowar, H. A., and Chapman, W. 2008a. "A review of journal policies for sharing research data," in *ELPUB*: Toronto Canada.

Piwowar, H. A., and Chapman, W. W. 2008b. "A review of journal policies for sharing research data," in *International Conference on Electronic Publishing (ELPUB2008)*.

Piwowar, H. A., and Chapman, W. W. 2010. "Public sharing of research datasets: A pilot study of associations," *Journal of Informetrics* (4:2) Apr, pp 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) Mar.

Plano Clark, V. L., and Creswell, J. W. 2008. *The Mixed Methods Reader*, Sage Publications: Thousand Oaks.

Polanyi, M. 1945. "The autonomy of science," *The Scientific Monthly* (60:2), pp 141-150.

Popper, K. R. 1968. *Conjectures and refutations: The growth of scientific knowledge*, Harper & Row: New York.

Posey, C., Lowry, P. B., Roberts, T. L., and Ellis, T. S. 2010. "Proposing the online community self-disclosure model: the case of working professionals in France and the UK who use online communities," *European Journal of Information Systems* (19:2) Apr, pp 181-195.

Powell, W. W. 1991. *Expanding the scope of institutional analysis*, University of Chicago Press: Chicago.

Powell, W. W., and Colyvas, J. A. 2008. "Microfoundations of institutional theory," in *The sage handbook of organizational institutionalism,* C. O. Greenwood, R. Suddaby and K. Sahlin (eds.), Sage: Thousand Oaks, CA.

Pryor, G. 2009. "Multi-scale data sharing in the life sciences: Some lessons for policy makers," *International Journal of Digital Curation* (4:3), pp 17-82.

Punch, K. F. 2005. *Introduction to social research: Quantitative and qualitative approaches*, Sage Publications Ltd.

Ragin, C. C. 1994. *Constructing Social Research: The Unity and Diversity of Method*, (Pine Forge Press: Thousand Oaks.

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

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

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

Reitsma, F., Laxton, J., Ballard, S., Kuhn, W., and Abdelmoty, A. 2009. "Semantics, ontologies and eScience for the geosciences," *Computers & Geosciences* (35:4) Apr, pp 706-709.

Ribes, D., and Lee, C. P. 2010. "Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories," *Computer Supported Cooperative Work-the Journal of Collaborative Computing* (19:3-4) Aug, pp 231-244.

Richardson, H. A., and 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), pp 561-589.

Richter, M. N. 1980. *The autonomy of science: An historical and comparative analysis*, Schenkman Pub. Co.: Cambridge, MA.

Roberts, J. 2000. "From know-how to show-how? Questioning the role of information and communication technologies in knowledge transfer," *Technology Analysis & Strategic Management* (12:4) Dec, pp 429-443.

Robinson, K. J. 2011. "The Rise of Choice in the US University and College: 1910-2005," *Sociological Forum* (26:3) Sep, pp 601-622.

Robson, G. S., Wholey, D. R., and Barefield, R. M. 1996. "Institutional determinants of individual mobility: Bringing the professions back in," *Academy of Management Journal* (39:2), pp 397-420.

Rogelberg, S. G., and Stanton, J. M. 2007. "Introduction understanding and dealing with organizational survey nonresponse," *Organizational Research Methods* (10:2), pp 195-209.

Roth, N. R., Sitkin, S. B., and House, A. 1994. "Stigma as a Determinant of Legalization," in *The Legalistic Organization,* S. B. Sitkin and R. J. Bies (eds.), Sage: Newbury Park, CA, pp. 137-168.

Rupidara, N. S., and McGraw, P. 2011. "The role of actors in configuring HR systems within multinational subsidiaries," *Human Resource Management Review* (21:3) Sep, pp 174-185.

Ryu, S., Ho, S. H., and Han, I. 2003. "Knowledge sharing behavior of physicians in hospitals," *Expert Systems with Applications* (25:1) Jul, pp 113-122.

Sacco, J. M., Scheu, C. R., Ryan, A. M., and 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), pp 852-865.

Saltz, J., Oster, S., Hastings, S., Langella, S., Kurc, T., Sanchez, W., Kher, M., Manisundaram, A., Shanbhag, K., and Covitz, P. 2006. "caGrid: Design and implementation of the core architecture of the cancer biomedical informatics grid," *Bioinformatics* (22:15), pp 1910-1916.

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

Scherbaum, C. A., and Ferreter, J. M. 2009. "Estimating statistical power and required sample sizes for organizational research using multilevel modeling," *Organizational Research Methods* (12:2), pp 347-367.

Schickore, J. 2008. "Doing Science, Writing Science," *Philosophy of Science* (75:3) Jul, pp 323-343.

Schutt, R. 2006. *Investigating the social world: The process and practice of research*, Pine Forge Press.

Schwartz, A., Pappas, C., and Sandlow, L. J. 2010. "Data Repositories for Medical Education Research: Issues and Recommendations," *Academic Medicine* (85:5) May, pp 837-843.

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

Scott, W. R. 1995. *Institutions and Organizations*, Sage Publications: Thousand Oaks, CA.

Scott, W. R. 2004. "Institutional Theory: Contributing to a Theoretical Research Program," in *Great Minds in Management,* K. G. Smith and M. A. Hitt (eds.), Oxford: New York, pp. 460-484.

Scott, W. R. 2007. *Institutions and organizations: Ideas and interests*, Sage Publications: Thousand Oaks, CA.

Sedberry, G. R., Fautin, D. G., Feldman, M., Fornwall, M. D., Goldstein, P., and Guralnick, R. P. 2011. "OBIS-USA A Data-Sharing Legacy of the Census of Marine Life," *Oceanography* (24:2) Jun, pp 166-173.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., and Mermelstein, R. J. 2012. "A Practical Guide to Calculating Cohen's f2, a Measure of Local Effect Size, from PROC MIXED," *Frontiers in Psychology* (3:111), pp 1-6.

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

Shin, D. H. 2008. "Understanding purchasing behaviors in a virtual economy: Consumer behavior involving virtual currency in Web 2.0 communities," *Interacting with Computers* (20:4-5), pp 433-446.

Shrout, P. E., and Fleiss, J. L. 1979. "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin* (86:2), pp 420-428.

Sitkin, S. B., and George, E. 2005. "Managerial trust-building through the use of legitimating formal and informal control mechanisms," *International Sociology* (20:3) Sep, pp 307-338.

Smith, D. H. 1981. "Altruism, volunteers, and volunteerism," *Journal of Voluntary Action Research* (10:1) 1981, pp 21-36.

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

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

Ssewamala, F. M., and Sherraden, M. 2004. "Integrating saving into microenterprise programs for the poor: Do institutions matter?," *Social Service Review* (78:3) Sep, pp 404-428.

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

Stein, L. D. 2008. "Wiki features and commenting - Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges," *Nature Reviews Genetics* (9:9) Sep, pp 678-688.

Steinhart, G. 2007. "DataStaR: An Institutional Approach to Research Data Curation," *IASSIST Quarterly*).

Sterling, T. D., and Weinkam, J. J. 1990. "Sharing scientific-data," *Communications of the ACM* (33:8) Aug, pp 112-119.

Strier, K. B., Altmann, J., Brockman, D. K., Bronikowski, A. M., Cords, M., Fedigan, L. M., Lapp, H., Liu, X. H., Morris, W. F., Pusey, A. E., Stoinski, T. S., and Alberts, S. C. 2010. "The Primate Life History Database: a unique shared ecological data resource," *Methods in Ecology and Evolution* (1:2) Jun, pp 199-211.

Suddaby, R. 2010. "Challenges for Institutional Theory," *Journal of Management Inquiry* (19:1) Mar, pp 14-20.

Swan, J. E., and Oliver, R. L. 1989. "Postpurchase communications by consumers," *Journal of Retailing* (65:4) Win, pp 516-533.

Szyliowicz, D., and Galvin, T. 2010. "Applying broader strokes: Extending institutional perspectives and agendas for international entrepreneurship research," *International Business Review* (19:4) Aug, pp 317-332.

Tabachnick, B. G., and Fidell, L. S. 2000. *Using Multivariate Statistics (4th ed.)*, Allyn and Bacon: Boston, MA.

Taylor, J. W. 1974. "Role of risk in consumer behavior," *Journal of Marketing* (38:2) 1974, pp 54-60.

Taylor, P. L. 2007. "Research sharing, ethics and public benefit," *Nature Biotechnology* (25:4) Apr, pp 398-401.

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

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., and Frame, M. 2011. "Data Sharing by Scientists: Practices and Perceptions," *Plos One* (6:6) Jun.

Teo, H. H., Wei, K. K., and Benbasat, I. 2003. "Predicting intention to adopt interorganizational linkages: An institutional perspective," *Mis Quarterly* (27:1) Mar, pp 19-49.

Thompson, R. L., Higgins, C. A., and Howell, J. M. 1991. "Personal computing: Toward a conceptual model of utilization," *MIS Quarterly: Management Information Systems* (15:1), pp 125-142.

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

Thornton, P. H., and Ocasio, W. 1999. "Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958-1990," *American Journal of Sociology* (105:3) Nov, pp 801-843.

Thornton, P. H., and Ocasio, W. 2008. "Institutional logics," in *The Sage Handbook of Organizational Institutionalism,* R. Greenwood, C. Oliver, R. Suddaby and K. Sahlin-Andersson (eds.), Sage: Thousand Oaks, CA.

Titah, R., and Barki, H. 2009. "Nonlinearities between Attitude and Subjective Norms in Information Technology Acceptance: A Negative Synergy?," *MIS Quarterly* (33:4), pp 827-844.

Tolbert, P. S. 1985. "Institutional Environments and Resource Dependence: Sources of Administrative Structure in Institutions of Higher Education," *Administrative Science Quarterly* (30:1), pp 1-13.

Tolbert, P. S., and 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), pp 22-39.

Uhlir, P. F. 2010. "Information gulags, intellectual straightjackets, and memory holes: Three principles to guide the preservation of scientific data," *Data Science Journal* (10), pp 1-5.

Van House, N. A., Butler, M. H., and Schiff, L. R. 1998. "Cooperative knowledge work and practices of trust: Sharing environmental planning data sets," *Proceedings of the ACM Conference on Computer Supported Cooperative Work*), pp 335-343.

Vandenabeele, W. 2007. "Toward a public administration theory of public service motivation - An institutional approach," *Public Management Review* (9:4) Dec, pp 545-556.

Venkatesh, V., and Davis, F. D. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46:2), p 186.

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), pp 425-478.

Vickers, A. J. 2006. "Whose data set is it anyway? Sharing raw data from randomized trials," *Trials* (7) May.

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

Wallis, J. C., Borgman, C. L., Mayernik, M. S., Pepe, A., Ramanathan, N., and Hansen, M. 2007. "Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries," pp. 380-391.

Wallis, J. C., Milojevic, S., Borgman, C. L., and Sandoval, W. A. 2006. "The special case of scientific data sharing with education," *Proceedings of the ASIST Annual Meeting* (43).

Walsh, J. P., and Hong, W. 2003. "Secrecy is increasing in step with competition," *Nature* (422:6934), pp 801-802.

Watson, S., and 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), pp 141-173.

Weber, N. M., Piwowar, H. A., and Vision, T. J. 2010. "Evaluating data citation and sharing policies in the environmental sciences," *Proceedings of the American Society for Information Science and Technology* (47:1), pp 1-2.

Weil, V., and Hollander, R. 1991. "Normative issues in data-sharing," in *Sharing Social Science Data: Advantages and Challenges,* J. Sieber (ed.), Sage: London, pp. 151-157.

Whitley, R. 2000. *The intellectual and social organization of the sciences*, Oxford University Press: Oxford.

Wicherts, J. M., Borsboom, D., Kats, J., and Molenaar, D. 2006. "The poor availability of psychological research data for reanalysis," *American Psychologist* (61:7) Oct, pp 726-728.

Wicks, D. 2001. "Institutionalized mindsets of invulnerability: Differentiated institutional fields and the antecedents of organizational crisis," *Organization Studies* (22:4), pp 659-692.

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

Wolins, L. 1962. "Responsibility for raw data," *American Psychologist* (17:9), pp 657-658.

Wong, E. M. L., and Li, S. C. 2008. "Framing ICT implementation in a context of educational change: A multilevel analysis," *School Effectiveness and School Improvement* (19:1), pp 99-120.

Wright, D. J., and Wang, S. W. 2011. "The emergence of spatial cyberinfrastructure," *Proceedings of the National Academy of Sciences of the United States of America* (108:14) Apr, pp 5488-5491.

Wu, W.-L., Lin, C.-H., Hsu, B.-F., and 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), pp 83-93.

Wu, W. Y., and Li, C. Y. 2007. "A contingency approach to incorporate human, emotional and social influence into a TAM for KM programs," *Journal of Information Science* (33:3), pp 275-297.

Zhang, Z., Zyphur, M. J., and Preacher, K. J. 2009. "Testing Multilevel Mediation Using Hierarchical Linear Models Problems and Solutions," *Organizational Research Methods* (12:4), pp 695-719.

Ziman, J. M. 2000. *Real Science: What It Is, and What It Means*, Cambridge University Press: Cambridge, U.K.; New York.

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), pp 5-16.

Zimmerman, A. S. 2003. *Data Sharing and Secondary Use of Scientific Data: Experiences of ecologists*, University of Michigan.

Zimmerman, A. S. 2008. "New knowledge from old data - The role of standards in the sharing and reuse of ecological data," *Science Technology & Human Values* (33:5) Sep, pp 631-652.

Zsidisin, G. A., Melnyk, S. A., and Ragatz, G. L. 2005. "An institutional theory perspective of business continuity planning for purchasing and supply management," *International Journal of Production Research* (43:16) Aug 15, pp 3401-3420.

Zucker, L. G. 1977. "Role of instituionalization in cultural persistence," *American Sociological Review* (42:5), pp 726-743.

Zucker, L. G. 1991. "The role of institutionalization in cultural persistence," in *The New Institutionalism in Organizational Analysis,* W. W. Powell and P. J. DiMaggio (eds.), University of Chicago Press: Chicago, IL, pp. 83-107.

Zucker, L. G., and Darby, M. R. 2004. "An Evolutionary Approach to Institutions and Social Construction: Process and Structure," in *Great Minds in Management,* K. G. Smith and M. A. Hitt (eds.), Oxford: New York, pp. 547-571.

Curriculum Vitae

YOUNGSEEK KIM

221 Hinds Hall
School of Information Studies
Syracuse University
Syracuse, NY 13244

Phone: (315) 464 – 0824
Fax: (315) 443 – 6886
E-Mail: ykim58@syr.edu

EDUCATION

Syracuse University (SU), School of Information Studies (iSchool)

Ph.D. in Information Science and Technology, August 2008 – June 2013

Dissertation Title: "Institutional and Individual Influences on Scientists' Data Sharing Behaviors"

Committee: Dr. Ping Zhang, Dr. Jeffrey Stanton, Dr. Kevin Crowston, & Dr. Jian Qin

M.S.in Information Management, August 2006 – August 2008

Seoul National University (SNU), March 1999 – August 2006, Seoul, Korea

B.A. in Religious Studies and Information & Multimedia Culture Studies

University of Missouri – Columbia (UMC), January – December 2005, Columbia, MO

Visiting Student at the School of Information Science & Learning Technologies

HONORS & AWARDS

Awards

[A.5] Best Paper Award, "Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure," by Youngseek Kim, Benjamin Addom, & Jeffrey Stanton, *International Digital Curation Conference*, Chicago, IL, December 6-8, 2010.

[A.4] Certificate in University Teaching in Recognition of Excellence in Professional Preparation for an Academic Career, Future Professoriate Program in the Graduate School at SU, 2010.

[A.3] Master's Prize in Information Management in Recognition of Excellence in Scholarship and Research, School of Information Studies at SU, 2008.

[A.2] Graduation with Top Honors (*Summa Cum Laude*), Seoul National University, 2006.

[A.1] Excellent Thesis Award, Writing Center at Seoul National University, 2006.
Thesis Title: "Korean Religious Communities' Changes Developed by the Internet Technology"

Scholarship & Fellowship

[S.4] Institute of Museum and Library Services (IMLS) eScience Fellowship, School of Information Studies at SU, 2011-2013.

[S.3] Jeffrey Katzer Doctoral Fellowship, School of Information Studies at SU, 2010-2011.

[S.2] The Korean Honor Scholarship, Embassy of the Republic of Korea in the U.S., 2010.

[S.1] Governmental Study Abroad Fellowship, Korean Culture and Content Agency, 2005.

Research Grants

[G.2] DataONE Usability and Assessment and Sociocultural Working Group Meeting, Knoxville, TN. May 1-3, 2012.

[G.1] Student Travel Support Grant from the ILMS and the Digital Curation Centre to attend *the International Digital Curation Conference*, Chicago, IL. December 2010.

PUBLICATIONS

Refereed Journal Publications

[J.3] **Kim, Y.**, & Stanton, J. M. (2012). Institutional and Individual Influences on Scientists' Data Sharing Practices. *Journal of Computational Science Education, 3(1)*, 47-56.

[J.2] **Kim, Y.**, Addom, B. K., & Stanton, J. M. (2011). Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure. *The International Journal of Digital Curation, 6(1)*, 125-138.

[J.1] Stanton, J. M., **Kim, Y.**, Oakleaf, M., Lankes, R. D., Gandel, P., Cogburn, D., & Liddy, E. D. (2011). Education for eScience Professionals: Job Analysis, Curriculum Guidance, and Program Consideration. *Journal of Education for Library and Information Science, 52(2)*, 79-94.

Refereed Conference Papers

[C.7] **Kim, Y.**, & Crowston, K. (2011). Technology Adoption and Use Theory Review for Studying Scientists' Continued Use of Cyberinfrastructure. Paper presented at *the Annual Meeting of the American Society for Information Science and Technology*, October 9-12, New Orleans, LA.

[C.6] **Kim, Y.**, Addom, B. K., & Stanton, J. M. (2010). Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure. Paper presented at *the International Digital Curation Conference*, December 6-8, Chicago, IL (**Best Paper Award**).

[C.5] Han, J., & **Kim, Y.** (2009). Obama Tweeting and Twitted: Sotomayor's Nomination and Health Care Reform. Paper presented at *the American Political Science Association Conference*, September 3-6, Toronto, Canada.

[C.4] **Kim, Y.**, Zhang, P., & Lankes, R. D. (2009). Future of Information Seeking on the Internet and Web Advertising: How Can We Guide Web Advertising for Users'

Information Seeking on a Website? Paper presented at *the iConference*, February 8-11, Chapel-Hill, NC.

[C.3] **Kim, Y.**, Howard, J., Ravindranath, S., & Park, J. S. (2008). Problem Analyses and Recommendations in U.S. DRM Security Policies. Paper presented at *the EuroISI*, December 3-5, Esbjerg, Denmark.

[C.2] Zhang, P., & **Kim, Y.** (2008a). Web Advertising: What Do We Know about its Acceptance and Impacts? A Meta-Analysis of the Literature. Paper presented at *the Pacific Asia Conference on Information Systems (PACIS)*, July 3-7, SuZhou, China.

[C.1] Zhang, P., & **Kim, Y.** (2008b). What Makes Web Advertisements Effective? Paper presented at *the China Summer Workshop on Information Management (CSWIM)*, June 29-30, Kunming, Yunnan, China.

Refereed Conference Posters

[P.6] **Kim, Y.** (2012). Factors Influencing STEM Researchers' Data Sharing Behaviors. Poster presented at *the Annual Meeting of the American Society for Information Science and Technology*, October 26-30, Baltimore, MD.

[P.5] Addom, B. K., **Kim, Y.**, & Stanton, J. M. (2011). eScience Professional Positions in the Job Market: A Content Analysis of Job Advertisements. Poster presented at *the iConference*, February 8-11, Seattle, WA.

[P.4] **Kim, Y.**, & Zhang, P. (2010). Continued Use of Technology: Combining Controlled and Automatic Processes. Poster presented at *the International Conference on Information Systems (ICIS)*, December 12-15, St. Louis, MO.

[P.3] **Kim, Y.**, Kim, M., & Kim, K. (2010). Factors Influencing the Adoption of Social Media in the Perspective of Information Needs. Poster presented at *the iConference*, February 3-6, Urbana-Champaign, IL.

[P.2] **Kim, Y.**, & Zhang, P. (2009). Individual Users' Adoption of Smart Phone Services. Poster presented at *the 8th Annual Workshop on HCI Research in MIS*, December 14, Phoenix, AZ.

[P.1] **Kim, Y.** (2009). How Does Web Advertising Affect Users' Information Seeking, Website Evaluation, and Source Evaluation? Poster presented at *the iConference*, February 8-11,
Chapel-Hill, NC.

Doctoral Colloquium

[D.1] Doctoral Seminar on Research and Career Development (2012). *The Annual Meeting of the American Society for Information Science and Technology*, October 26-30, Baltimore, MD.

Technical Reports and White Papers

[T.2] Crowston, K., & **Kim, Y.** (2011). Scientists' Adoption of New Technologies. White paper submitted to the Data Observation Network for Earth (DataONE) project.

[T.1] Mueller, M., & **Kim, Y.** (2009). Economic Factors in the Allocation of IP Addresses. Report submitted to the International Telecommunication Union.

Invited Talks and Presentations

[I.3] "Institutional and Individual Influences on Scientists' Data Sharing Practices" Presentation given at the Graduate School of Convergence Science & Technology, Seoul National University,
June 29, 2012.

[I.2] "Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure" Presentation given at the eScience Fellow Student Meeting in the School of Information Studies, Syracuse University, February 16, 2011.

[I.1] "Mobile Computing Technology Engagement: Combining Adoption and Continued Use Theories" Presentation given at the Poster Reception with Board of Advisors in the School of Information Studies, Syracuse University, May 7, 2010.

RESEARCH EXPERIENCE

School of Information Studies at SU

Funded Research Projects

[R.4] **Scientific Data Sharing and Reuse Project** (*Supported by ILMS eScience Fellowship*)
Research Fellow, August 2011 – Present, Syracuse, NY

- Examine scientists' data sharing and reuse practices across the entire data life cycle
- Conducted interviews with STEM researchers to understand their data sharing and reuse practices and analyzed the interview transcriptions by using content analysis methods
- Reviewed literature in data curation, scholarly communication, and knowledge management

[R.3] **Data Observation Network for Earth (DataONE) Project**
(*Sociocultural Issues Working Group Member: Dr. Kevin Crowston, NSF funded*)
Research Assistant, February 2011 – May 2011, Syracuse, NY

- Conducted literature review in STEM researchers' adoption and use of cyberinfrastructure
- Developed a theoretical framework for studying scientists' adoption and use of cyberinfrastructure and wrote a white paper for the DataONE project

[R.2] **Cyber-Infrastructure Facilitators Project** (*PI: Dr. Jeffrey Stanton, NSF funded*)
Research Assistant, August 2009 – August 2010, Syracuse, NY

- Conducted literature review in eScience and IT workforce development under STEM
- Analyzed qualitative data from interviews and focus groups by using content analysis methods
- Initiated writing two papers on job analysis of eScience professionals with the project results

[R.1] **IPv6 Address Allocation and Deployment Project**
(*PI: Dr. Milton Mueller, International Telecommunication Union funded*)
Research Assistant, May 2009 – August 2009, Syracuse, NY

- Collected the IPv6 allocation data in each country and wrote a summary report
- Conducted literature review in IPv6 deployment and interviewed experts in the domain area

Contributions to Grant Writing

[E.2] **Scientific Computing to Illuminate Dark Scientific Data** (*PI: Dr. Kevin Crowston*)
Submitted to the Expeditions in Computing Program at NSF, March 2012 (*Not decided*)

- Provided an overview of scientists' data practices focusing on data sharing and reuse
- Identified and introduced two senior scientists to the project team as senior personnel

[E.1] **Microblogging and Health Care Reform** (*PI: Dr. Jongwoo Han*)
Submitted to the Political Science Program at NSF, August 2010 (*Not funded*)

- Conducted a preliminary study on Tweet data regarding President Obama's health care reform by collecting Tweet data through API and analyzing the contents of Tweets
- Wrote a part of research method section focusing on data collection and opinion analysis

Information & Multimedia Culture Studies (IMCS) at SNU

[O.3] **Web Portal Study***, Co-team project by NHN Corporations and IMCS at SNU*
Research Assistant, May 2007 – August 2007, Syracuse, NY

- Performed literature review on the Web portals and surveyed portals in the U.S. market
- Presented the current Web portal trends and suggested a future roadmap for NHN Corp

[O.2] **Global Leadership Academy for Cultural Industry**, *IMCS at SNU*
Program Assistant, March 2006 – June 2006, Seoul, Korea

- Developed course curriculum regarding cultural industry by surveying cases in the U.S.
- Served faculty members to create and review their courseworks and case studies

[O.1] Research Assistant, June 2003 – December 2003, Seoul Korea
Culture Technology Implementation for Korea Telecom (KT), *IMCS at SNU*

- Surveyed the content market of telecommunication companies in the world
- Mapped out and suggested available information content services for KT

Universal Design Exhibition, *IMCS at SNU*

- Researched the user-interface problems of extant electronic appliances
- Presented the user-interface prototype of electronic appliances for universal design

Next-Generation Communication Platform Study, *Co-team project by Daum and IMCS*

- Conducted literature review on the communication environment and technology
- Presented oncoming Internet technologies for Daum's service innovation

TEACHING EXPERIENCE

School of Information Studies at SU

Instructor

IST449 **Human Computer Interaction** (Spring 2011, Spring 2012, and Summer 2012*)

This course covers the design, evaluation, and implementation of interactive computing systems for human use. It introduces theories of human psychology, principles of computer systems and user interface designs, a methodology of developing effective HCI for information systems, and diverse methods involved in evaluations. (*This course was taught with IST649 in summer 2012.)

IST649 **Human Interaction with Computers** (Summer 2012–*Online Format*)

This course is the graduate version of IST449 *Human Computer Interaction*. This course covers similar content as IST449, plus the multi- and inter-disciplinary nature of HCI and various HCI issues in the organizational and societal contexts.

IST614 **Management Principles for Information Professionals** (Summer 2011–*Online Format*)

This course is a required course for both the M.S. in Library and Information Science and the M.S. in Information Management. It introduces graduate students to the profession and practice of management in the information field. It is designed to illustrate management themes which are common to most organizational contexts, and it covers the theoretical concepts of organization theory, and managerial principles and techniques.

IST621 **Introduction to Information Management** (Fall 2010 and 2011)

This course is the gateway course for the M.S. in Information Management program. It covers the issues and challenges involved in managing information resources in organizations. The discussion session is designed for students to participate in weekly discussions pertaining to the topics covered each week in IST 621.

Co-Instructor

IST619 **Applied Economics for Information Managers** (Spring 2012–*Online Format*)

Lead Instructor: Prof. Ian MacInnes

- Modified fourteen-week course lectures and relevant assignments in an online format
- Created two exams, consulted on weekly quizzes, and led online discussions on the Blackboard Learning System

Guest Lecturer

IST654 **Information Systems Analysis** (February 22 and September 28, 2011)

Instructor: Prof. Kevin Crowston

- Lectured on Human-Centered System Development Life-Cycle and Interface Design
- Reviewed and provided feedback on students' project proposals in information system designs

Teaching Assistant

IST755 **Strategic Management of Information Resources** (Spring 2011)

Instructors: Prof. Michelle Kaarst-Brown and Prof. Herbert Brinberg (2 sections)

- Graded and provided extensive feedback for assignments and semester-long projects
- Held office hours to consult with students on their semester-long capstone projects

IST195 **Information Technology** (Fall 2008 and Spring 2009)

Instructor: Prof. Jeffrey Rubin

- Redesigned IT laboratory materials and supported lab sessions for undergraduate students
- Assisted the professor with class lecture preparation and graded weekly labs and final projects

IST466 **Professional Issues in Information Management & Technologies** (Fall 2007)

Instructor: Prof. Murali Venkatesh

- Supported the professor and students technically in using Adobe Premier and Encore
- Assisted students with their class projects in regard to community wireless service

Teaching Practica

IST649 **Human Interaction with Computers** (Fall 2009), Prof. Ping Zhang

- Updated the reading list in the syllabus and created two short assignments
- Developed a lab module on system evaluation and graded, and provided feedback to students

IST619 **Applied Economics for Information Managers** (Spring 2009), Prof. Ian MacInnes

- Developed exams on economics and graded and provided extensive feedback for the exams
- Observed online class management and researched teaching techniques in online courses

MIS325 **Introduction to Information Systems for Managers** (Fall 2008–*School of Management*),

Prof. Joseph Treglia (Instructor) and Prof. Murali Venkatesh (Faculty Advisor)

- Created and provided two lectures about website design and development in two sessions
- Provided consultations for students' questions and problems in their projects in two sessions

IST623 **Introduction to Information Security** (Summer 2008), Prof. Joon Park

- Created three lectures on social engineering, privacy, and legal issues in security
- Developed a lab module regarding information privacy and tested it with students

WORK EXPERIENCE

[W.5] **Project Manager**, August 2010 – December 2011, Syracuse, NY

Korean War Veterans' Digital Memorial (KWVDM) Project in the Maxwell School at SU

- Wrote a funding proposal with Prof. Jongwoo Han to be submitted to the Ministry of Patriots and Veterans Affairs of the Republic of Korea
- Supervised 4 graduate students and 2 professionals (a Web designer and a Web developer) to design and develop the KWVDM website (http:www.kwvdm.org)

[W.4] **Web Master**, August 2006 – August 2008**,** Syracuse, NY
Entrepreneurship and Emerging Enterprises (EEE) in the School of Management at SU

- Maintained the 170 Web pages of the EEE website and managed the Web databases of EEE
- Designed and created websites for 2008 USASBE Conference and Syracuse Women Business Center with HTML, CSS, JavaScript, and ASP
- Redesigned WISE Symposium and South Side Innovation Center (SSIC) websites and developed new online registration and payment system with ASP, and Access

[W.3] **Assistant Technology Coordinator**, August 2005 – December 2005, Columbia, MO
School of Information Science & Learning Technologies at UMC

- Assisted undergraduates in using Web conference software by creating a 40-page guide
- Coordinated Web conferences with China and maintained the equipment

[W.2] **Assistant Web Developer**, February 2005 – November 2005, Columbia, MO
School of Information Science & Learning Technologies at UMC

- Joined in the collaborative learning for K-12 students in Missouri with other countries
- Involved in developing the portions of 40 Web pages for the Show Me the World website

[W.1] **Sergeant and Team Leader**, *The Korea Army*, January 2000 – March 2002, Cheongwon, Korea

- Managed the reserve force management system by updating reserve force data
- Developed annual training plans for reserve forces and participated in field trainings with them

PROFESSIONAL INVOLVEMENT

Academic and Professional Memberships

- American Society of Information Science and Technology
- Association for Information Systems
- Association for Library and Information Science Education

Journal Article Reviewing

- Journal of the Association for Information Systems (2009-2012)
- Communications of the Association for Information Systems (2010-2011)
- AIS Transactions on Human-Computer Interaction (2010-2011)
- Journal of Computer-Mediated Communication (2010)

Conference Reviewing

- International Conference on Information Systems (2008-2012)
- Americas Conference on Information Systems (2010, 2012)
- Hawaii International Conference on System Sciences (2011)
- ICIS Pre-Workshop on HCI Research in MIS (2009-2011)
- Mediterranean Conference on Information Systems (2010)
- iConference (2009-2010)
- Symposium on Access Control Models and Technologies (2009)

University Service

- Faculty Search Committee, School of Information Studies at SU (2011-2012)
- Data Science Certificate of Advanced Study (CAS) Development Initiative, School of Information Studies at SU (2009-2010)
- University Assessment Council, School of Information Studies at SU (2009-2010)
- Undergraduate Committee, School of Information Studies at SU (2008-2009)

Community Service

- Volunteer, American Society for Information Science and Technology Annual Meeting (2012)
- Board Member, Korean Student Association at Syracuse University (2009-2011)
  - Organized the Korean Film Festival and the Korean Board Game Night at SU
- Web Master, Central New York Korean School (CNYKS) (2007-2009)
  - Created the CNYKS website (http://www.cnyks.org) and updated it regularly
- Instructor, South Side Innovation Center (2007-2009)
  - Taught Microsoft Excel, Access, and Web design for local business owners and employees
- Technology Coordinator, Central New York InterFaith Works (2007)
  - Supported Recruitment and Retention: TOOLS for a Diverse Workforce Conference