FACULTY SELF-ARCHIVING BEHAVIOR:

FACTORS AFFECTING THE DECISION TO SELF-ARCHIVE

by

Ji-Hyun Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Elizabeth Yakel, Chair
Professor Paul N. Courant
Associate Professor Margaret L. Hedstrom
Assistant Professor Soo Young Rieh

To my mother

ACKNOWLEDGEMENTS

During my dissertation research, I have been helped tremendously by a number of people who give me intellectual and emotional support. I am delighted to take this opportunity to show my sincere gratitude toward their support and care.

First of all, I would like to thank all my committee members, especially my Advisor and Chair, Dr. Elizabeth Yakel. Her guidance, encouragement, and feedback were crucial for me to start and conduct the dissertation research. She has always been interested in my progress and promptly offered advice whenever I needed. Her advice helped me focus on aspects of importance and move forward to complete the study. Besides my dissertation, she is the one who had led me to various research activities – research projects, conference presentations, and publications - which have helped me to learn research skills and grow to be a researcher. She has been a wonderful advisor, with the patience and trust to nurture and motivate me to continue my development in Ph.D. studies.

Dr. Margaret Hedstrom has provided insightful comments about the study, especially when developing my survey questionnaire, as well as for the final revision of my dissertation, which makes clear the significance of the study. She also allowed me to join the University of Michigan Institutional Repository (Deep Blue) meeting as an observer for a year. This was a great opportunity for me to learn the administrative aspects of the repository and eventually helped me write the practical implications of my study.

Dr. Soo Young Rieh has been willing to discuss her opinions about my study, which helped it become more focused and organized. Her comments for the final revision

were detailed and substantive, with excellent suggestions for changes and improvements. In addition to research discussions, she has taken care of my personal situation and shared her experiences with me. Her advice was greatly helpful for managing time and life – particularly during the writing of my dissertation.

Dr. Paul Courant was a great outside member, who provided penetrating comments about my study in the proposal defense, to refocus the study on general self-archiving behavior, rather than only IR contribution. He also made substantial comments on my survey questionnaire, which helped me collect meaningful data.

In addition to my committee members, I would like to thank Dr. Stephanie Teasley, for her understanding and advice when I was expecting and planning to go on detached study status. I also want to thank SI support staff, especially Sue Schuon, Ann Verhey-Henke, and Debbie Apsley. Sue has managed all the administrative and logistic issues while I mailed invitation letters and prepared my oral defense. Ann helped figure out how to distribute cash incentives to faculty members outside the University of Michigan, and Debbie Apsley has done excellent work on translating interviews in an efficient and accurate manner.

I also want to extend my appreciation to my SI colleagues for their help, comments, and encouragement. Airong Luo has been an especially good friend of mine in SI, always willing to listen and give feedback. I also wish to thank John Lin, Dana Walker, Nikhil Sharmar, Erik Johnston, Beth St. Jean, Trond Jacobsen, Jinfang Niu, Magia Krause, Yong-Mi Kim, Jina Huh, and Il-Hwan Kim. In particular, Il-Hwan greatly helped me set up the online survey questionnaire, corresponding to paper invitation letters.

Finally, I would like to extend special thanks to my family members, who always wanted me to succeed in the Ph.D. studies. My mom is the one who introduced the field of Library and Information Science to me and encouraged me to continue to study in the U.S. Without her support and love, I would not have even considered starting the doctoral

studies. Even though my father passed away 14 years ago, he would be very pleased with this achievement. My aunt has also given substantial support and encouragement to me. My older sister (Soo Hyeon), brother-in-law (Seung-Ryong), niece (Hye Joo), younger brother (Sung Hwan), and sister-in-law (Myung Hee) always encouraged me during my studies. My mother-in-law came from Korea and helped with all of the house chores when I was busiest for preparing my oral defense, and I thank her for that. At last, I am very grateful for my husband (Ho-Chul) and my daughter (Chloe) who were each motivators for me to complete my dissertation. I really appreciate my husband's understanding and patience during the time of my dissertation writing. I also wish to thank Chloe's baby sitter (Alexis) for her tender-loving care for Chloe, which helped me focus on writing my dissertation.

TABLE OF CONTENTS

DEDICATION ..... **ii**

ACKNOWLEDGEMENTS ..... **iii**

LIST OF FIGURES ..... **x**

LIST OF TABLES ..... **xi**

LIST OF APPENDICES ..... **xiii**

ABSTRACT ..... **xiv**

CHAPTER 1 INTRODUCTION **1**

Definitions of Self-archiving 4

Forums for Self-archiving 5

Self-archived Research Materials 7

Motivations of the Study 9

Conceptual Framework 12

Socio-Technical Networks 13

Social Exchange Theory 16

Research Design of the Study 24

Significance of the Study 26

CHAPTER 2 LITERATURE REVIEW **27**

Disciplinary Variations in Scholarly Electronic Publishing 30

Electronic Journals 34

Definitions 34

Factors affecting E-journal Submission 36

Factors affecting OA Journal Submission 39
Self-archiving Practices 43
Disciplinary Repositories in Science 43
Working Paper Series: Guild Publishing Model 47
Institutional Repositories 49
Factors affecting Self-archiving Behavior 53
Cost Factors 54
Extrinsic Benefits 57
Intrinsic Benefit 63
Contextual Factors 65
Individual Traits 68
CHAPTER 3 METHODOLOGY 70
Target Population and Sampling Strategies 71
Survey Instrument Development 74
Measurement of Dependent Variables 74
Measurement of Independent Variables 75
Pilot Survey 79
Validation 80
Survey Data Collection 82
Survey Sample vs. Respondents 83
Interview Data Collection 89
CHAPTER 4 RESULTS 91
Survey Data Analysis 92
Self-archiving Practices 92
Factors affecting Self-archiving 109
Summary of Survey Data Analysis 120

Interview Data Analysis 121
Research Materials 122
Teaching Materials 133
Venues for Self-archiving 135
Factors affecting Self-archiving 147
Summary of Interview Data Analysis 191
CHAPTER 5 DISCUSSION 193
Self-Archiving Practices 195
Importance of Peer Review 195
Self-archiving Non-Textual Scholarly Content in IRs 198
Personal Web Pages: Control and Flexibility 201
IRs: Long-term Accessibility and Copyright Management 202
Motivating Factors 209
Altruism: Support for the Spirit of Open Access 209
Positive or Neutral Impact on Tenure and Promotion 210
Self-archiving Culture and Peer Pressure 211
Technical Skills, Age, and Rank 212
Impeding Factors 213
Confusion over Copyright Issues 213
Concerns about Additional Time and Effort 214
Fear of Plagiarism 215
CHAPTER 6 CONCLUSIONS 217
Theoretical Implications 217
STIN Model 217
Social Exchange Theory 219
Practical Implications 220

Limitations 222

Directions for Future Research 222

APPENDICES 226

BIBLIOGRAPHY 247

LIST OF FIGURES

Figure 1.1. Model of Factors affecting Self-archiving Behavior..... 23

Figure 3.1. Number of titles in each university's IR (in February 2006)..... 88

Figure 4.1. Percentages of Research Papers and Data sets Self-archived 96

Figure 4.2. Other Content Types Self-archived 97

Figure 4.3. Frequency of Self-archiving on Personal Web Pages..... 99

Figure 4.4. Frequency of Self-archiving on Research Group/Center/Lab Web Sites 100

Figure 4.5. Frequency of Self-archiving on Departmental Web Sites..... 101

Figure 4.6. Frequency of Self-archiving on Disciplinary Repositories 102

Figure 4.7. Frequency of Self-archiving on Institutional Repositories (IRs)..... 102

Figure 5.1. Updated Model of Factors affecting Faculty Self-archiving..... 216

LIST OF TABLES

Table 2.1. Factors affecting self-archiving identified from literature 69

Table 3.1. Independent Variables and Corresponding Hypotheses 70

Table 3.2. Operationalized Definitions of Costs, Benefits and Contextual Factors .. 75

Table 3.3. Concept Map 77

Table 3.4. Dates of Distributing Survey, Follow-up E-mails, and Closing the Survey 83

Table 3.5. Twelve Disciplines of IR non-contributor Sample 84

Table 3.6. Sample by Rank 85

Table 3.7. Respondents by Rank 85

Table 3.8. Sample by Disciplines 86

Table 3.9. Respondents by Disciplines 86

Table 3.10. Sample by University 88

Table 3.11. Respondents by University 89

Table 4.1. Self-archivers by Rank 92

Table 4.2. Self-archivers by Disciplines 92

Table 4.3. Academic Rank of Respondents 93

Table 4.4. Proportion of Self-archivers to Total Respondents by Disciplines 93

Table 4.5. How Long Respondents Had Self-archived 94

Table 4.6. People who are Actually Involved in Self-archiving 94

Table 4.7. Research Papers and Book Chapters Posted by Self-archivers 95

Table 4.8. Venues used by Self-archivers..... 98
Table 4.9. Awareness of IRs..... 104
Table 4.10. How respondents learned about the IR..... 104
Table 4.11. Real IR Contributors by Rank..... 105
Table 4.12. Real IR Contributors by Discipline..... 106
Table 4.13. Real IR Contributors by University..... 106
Table 4.14. Current Self-archiving vs. Future self-archiving..... 107
Table 4.16. Motivating factors affecting IR contribution in the Future..... 108
Table 4.17. Reliability of Factors..... 109
Table 4.18. Descriptive Statistics of Total Scores by Factors..... 110
Table 4.19. Self-reported Technical Skills..... 111
Table 4.21. Dependent Variables..... 112
Table 4.22. Independent Variables used for the Regression Analyses..... 112
Table 4.23. Results of OLS Regression..... 113
Table 4.24. Average Scores on Self-archiving Culture by Discipline..... 116
Table 4.25. Average Scores on Technical Skills by Discipline..... 117
Table 4.26. Results of Logistic Regression Analysis regarding IR Contribution... 118
Table 4.27. Interviewees by Rank..... 121
Table 4.28. Interviewees by Discipline..... 121
Table 4.29. Significant factors identified in the OLS regression analysis..... 147

LIST OF APPENDICES

Appendix A. Final Survey Instrument.....227
Appendix B. Invitation Letter..... 239
Appendix C. Interview Protocols..... 240
Appendix D. Statistical Methods 242
Appendix E. Disciplines of Survey Respondents..... 245

ABSTRACT

A transformation in scholarly communication is occurring due to the interactions among Internet technologies, new ways of accessing and disseminating scholarly content, as well as changes in the legal, economic, and policy aspects of scholarly publication systems. Self-archiving – the placement of research material on publicly accessible web sites – is an emerging practice used to disseminate scholarly content in a cost-effective and timely manner. This practice is supported by university libraries and public funding agencies through the support or provision of Open Access repository services. Nevertheless, many repositories suffer from low rates of participation. Institutional Repositories (IRs), in particular, have difficulty recruiting content from faculty members whose conduct research and generate a wide variety of research materials. To address this problem, I investigate the motivational factors affecting faculty to participation in various forms of self-archiving practices.

Based on the socio-technical network framework, this study views self-archiving practices as intertwined with technologies and social factors. The factors identified include cost, benefit, and contextual aspects of self-archiving, in addition to individual characteristics. To examine these significant factors affecting self-archiving, my research design involves triangulation of survey and interview data of faculty members sampled from 17 Carnegie Research Universities with DSpace IRs. The sample is also stratified by academic discipline due to existing evidence of variation based on fields.

The analysis of survey responses from 684 professors and 41 phone interviews found that the factor of altruism has the strongest effect on faculty self-archiving. This factor, however, is characterized more by reciprocity, rather than pure altruism. Self-archiving culture has the second greatest impact on the decision to self-archive. Therefore, faculty self-archiving is influenced greatly by intrinsic benefits or disciplinary norms, as opposed to extrinsic benefits. Concerning IRs in particular, results shows that the primary reason professors contribute to the repositories is the perceived ability of IRs to preserve scholarly content. This implies that digital preservation should be significantly more a core function of IRs. IR contributors are also concerned about copyright than non-contributors. Thus IR staff need to provide guidance for copyright management to alleviate this concern and any confusion.

CHAPTER 1

INTRODUCTION

Prior to the widespread use of computers and the Internet, scholarly communication was viewed as a 'social system' of which a major activity was the interaction of scholars, and its key elements were social institutions whose internal social processes responded to events of the system. These interactions were largely based on "informal" communication, information-exchange activities and "formal" communication, peer-reviewed articles and books (Garvey and Griffith, 1967). This clear distinction between informal and formal communication seemed reasonable; considering only print publication channels, monitored with discipline-specific standards, existed. Garvey and Griffith also emphasized some degree of orderliness in scholarly communication because of the relative constancies of scholars' goals of dissemination, as well as relatively stable social norms.

The perspective on scholarly communication systems suggested by Garvey and Griffith has been altered given the advent of electronic media and networked information. Lynch (1992) argued that while information technology changed existing assumptions of scholarly communication patterns, fully understanding these changes was difficult. He pointed out that although some elements of traditional paper-based publishing systems would be employed in the digital environment, it was not essential that the classical view of scholarly publishing be completely applied to the new situation. Thus, Lynch claimed that it did not matter whether various forms of creating, disseminating and communicating digital information were classified as scholarly publishing or not. These

processes would "eventually supplant publishing within the system of scholarly communication to the extent that they more effectively met the needs previously fulfilled by traditional paper-based publishing" (p.6).

Kling and Lamb (1996) also investigated technological impact on scholarly publishing. They found numerous non-empirical studies claiming that the use of information technology for scholarly publishing brought either optimistic or pessimistic consequences. The optimistic view termed 'technological utopianism' emphasized achieving social progress but did not consider conflict or competition. For example, one utopian argument was that electronic publishing enabled direct communication between authors and readers by cutting out mediators, such as publishers and libraries. In contrast, the pessimistic view referred to as 'technological anti-utopianism' cited various concerns about the technological effects on scholarship, such as the demise of libraries, unequal access to digital information, and difficulties in controlling the quality of scholarly content. Both perspectives presented simple causality between information technology and its consequences, and did not adequately explain the interactions between technology, people and their work practices.

As an alternative vision of scholarly electronic communication, Kling et al. (2003) analyzed various electronic Scholarly Communication Forums (e-SCFs), such as electronic journals or collaboratories, as 'socio-technical networks'. These networks referred to collections of social groups and artifacts incorporated in social and technological relationships. From this socio-technical perspective, social and technological aspects were neither meaningfully separable nor reducible to one another. The use of e-SCFs was shaped through interactions within and between social groups and a set of technologies.

Using this socio-technical approach as a general framework, this study investigates university faculty members' self-archiving behavior and their motivations and concerns about such activities. Self-archiving refers to depositing scholarly content in

publicly accessible websites, which range from personal and departmental web pages to disciplinary or institutional repositories. Scholarly content deposited in those forums also varies in versions and formats. While the concept of self-archiving is not clearly defined with respect to forums and content, there is a consensus on the role of self-archiving in complementing the current scholarly publishing system, as opposed to substituting for it (Crow, 2002; Guédon, 2003; Lynch, 2003; Swan and Brown, 2005). Faculty members create various digital documents and objects in the process of research, and publish some of those through the normal scholarly publishing channels, such as peer-reviewed journals, books or conference proceedings. Prior or subsequent to publication, the research output may also be made publicly accessible on the Internet. This self-archiving practice has been accomplished by individual or groups of researchers through their preferred web sites. For instance, personal web pages have been found to be the most commonly used forum for self-archiving (Gadd et al., 2003a; Swan and Brown, 2005). In addition, repositories of electronic preprints, technical reports or working papers have been actively used in some disciplines, for example, physics, mathematics, and computer science. Most recently, the Institutional Repository (IR) has emerged as a new method for self-archiving with emphasis on long-term and stable accessibility of scholarly materials created within a university community. This institution-based forum, however, has not yet been widely adopted by faculty members (Kim, 2006).

This study examines the wide range of faculty members' self-archiving behavior, which resides in the blurred area between informal and formal scholarly communication. Several empirical studies examined academic authors' self-archiving behavior quantitatively (Allen, 2005; Gadd et al., 2003a; Lawal, 2002; Pelizzari, 2003; Swan and Brown, 2005) and qualitatively (Davis and Connolly, 2007; Foster and Gibbons, 2005; Van House, 2003). However, faculty members' motivations for depositing their materials on publicly accessible websites, and how they make such decisions as what versions of materials to deposited and where to place those materials are not known. Examining

motivational factors that influence the decision to self-archive contributes to the body of literature regarding the transformation of scholarly communication, as well as to a better understanding of the practices of disciplinary and institutional repositories.

The following sections present (1) definitions of self-archiving; (2) motivations of the study and research questions; (3) conceptual framework; (4) research design; and (5) significance of the study.

Definitions of Self-archiving

A broad definition of self-archiving is "[depositing] a digital document in a publicly accessible website, preferably an OAI [Open Archives Initiative]-compliant Eprint Archive" (BOAI, 2006). In fact, the scope of digital documents and publicly accessible websites differ across the studies regarding self-archiving. In the context of self-archiving, digital documents usually refer to research literature including a copy of refereed, published articles, and pre-refereed drafts (Harnad, 2001; Swan and Brown, 2005), although other types of scholarly materials can be self-archived. In particular, Lynch (2003) and Crow (2002) suggested that IRs include not only electronic preprints and postprints, but also a wide range of teaching and research materials, such as technical reports, data sets, and art work. The present study focuses on several types of digital research materials including refereed, published articles, pre-refereed drafts, unrefereed papers (technical reports, working papers or project reports), book chapters, and data sets. These types of research work are created across almost all disciplines.

In surveying self-archiving forums, Coleman and Roback (2005) focused only on OAI-compliant repositories, whereas Swan and Brown (2005) included personal web pages as a self-archiving forum, in addition to OAI-compliant repositories. Documents in such repositories are tagged with XML-based metadata, which are harvested through OAI-PMH (Protocol for Metadata Harvesting). This mechanism allows documents in

individual OAI-compliant repositories to be accessible through Internet search engines, and seemingly those documents are placed in one huge repository. OAI compliance is a desirable feature to improve accessibility and publicity or items in open access repositories, but not a necessary condition for self-archiving (Harnad, 2004). In addition, only a minority of researchers currently self-archive their work in OAI-compliant repositories including IRs and some disciplinary repositories (Antelman, 2006). The present study, thus examines a broad range of forums that encompass both OAI-compliant repositories and other publicly accessible websites that faculty members use for self-archiving. Those forums are described in the following section.

Forums for Self-archiving

IRs and disciplinary repositories are regarded as preferable forums for self-archiving because they provide features and services that enhance access to scholarly content. In particular, arXiv, the disciplinary repository for the physics community, is considered a successful model of self-archiving. It was initiated at the Los Alamos National Laboratory in 1991, accepting electronic preprints in high energy physics. Currently, the Cornell University Library implements arXiv, which includes around 370,000 articles in physics, mathematics, and computer science (ArXiv.org, 2006).

The wide adoption of arXiv is attributed to the fact that it satisfies authors' needs for early dissemination of research work. Most pre-prints in arXiv are eventually published in scientific journals. The arXiv case demonstrates how research can be validated at two levels – open exchange of arXiv pre-prints, which later became refereed and published articles counted for tenure and promotion (Gunnarsdóttir, 2005). arXiv also provides features for formatting TeX documents into web-friendly output, such as HTML or PDF. Scientists commonly use TeX to represent mathematical formulae and other complex models. Therefore, the formatting function offers a great advantage for

authors submitting their TeX documents in arXiv (Krichel and Warner, 2001). Inspired by the success of arXiv, several disciplinary repositories have emerged in other disciplines, although they have not been as successful as arXiv. CogPrints in cognitive science, RePEc in economics, and DLIST in Library and Information Science are examples in operation at this time.

These emerging self-archiving practices and the OAI-supported interoperability of individual repositories drove the social and technical impetuses to develop IRs. IRs are defined as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members" (Lynch, 2003, p.2). In this definition, a major service of IRs becomes information management over time and the migration of digital content. This makes preservation of scholarly content an essential role of IRs. Documents in separate IRs can also be made accessible through major search engines because some repositories are OAI-compliant.

An increasing number of research universities in the U.S. have implemented or plan to implement an IR. Lynch and Lippincott (2005) found that out of 97 universities categorized as Carnegie 'doctoral universities', 40% already operated IRs. Among non-implementers, 88% were found to be in the planning stage of IR implementation. Similarly, the nationwide census of IRs in the U.S. conducted by the MIRACLE (Making Institutional Repositories a Collaborative Learning Environment) research project, also found that a vast majority of institutions that have implemented or pilot-tested IRs, were Carnegie doctorate-granting universities (Markey et al., 2007). Those findings indicate that IRs are becoming a component of the technical infrastructure in doctoral research institutions. Potential contributors to IRs include faculty, students and staff in universities, whereas faculty members are considered the crucial contributors of scholarly content to IRs. However, several studies noted that it has been difficult to get faculty

members to contribute (Chan, 2004; Davis and Connolly, 2007; Foster and Gibbons, 2005; Markey et al., 2007; Pelizzari, 2005; Young, 2002).

Despite the emergence of the more 'preferable' disciplinary and institutional repositories, academic authors have made more frequent use of personal web pages for self-archiving (Gadd et al., 2003a; Swan and Brown, 2005; Wojciechowska, 2007). Wojciechowska's (2007) surveyed the self-archiving behavior of mathematicians and computer scientists in France. She found that 66% of the mathematicians and 63% of the computer scientists (n=128) had self-archived pre-prints and post-prints on their personal websites. Personal websites were also found to be a popular venue from which users obtained OA articles. Miller (2006) also pointed out that 65% of her survey respondents had used self-archived articles from personal websites, whereas only 43% had used articles from either disciplinary or institutional repositories. Gadd et al. (2003a) also identified other types of websites, such as research group or project websites, co-authors' websites, and archives run by professional bodies as venues for self-archiving. Moreover, Kling et al. (2002) suggested websites of technical reports or working paper series supported by academic departments or schools were also used. The aforementioned forums contain various kinds of research/teaching materials posted by academic authors. Since research materials are the primary interest of the present study, the next section describes types of research work self-archived in those forums.

Self-archived Research Materials

Disciplinary repositories are usually termed 'e-print repositories', although the notion of 'e-prints' is not clearly defined. Some studies defined 'e-prints' as electronic versions of preprints (Brown, 2001; Lawal, 2002). The definition of preprints, however, varies. Garvey's (1979) definition of a preprint was "a prepublication draft of the manuscript which is submitted to a journal". Based on this notion, Kling (2004) defined a

preprint in a strict sense as "articles that have been accepted for a specific venue" (p. 600). Using this definition, Kling suggested that articles in disciplinary repositories were not electronic 'preprints' because they were not accepted by journals. Instead, Kling termed those articles as 'electronic manuscripts (e-scripts)', which referred to documents that are to be submitted and/or are under review. Yet, Guédon (2003) and Pinfield (2003) named the documents under review as 'pre-refereed' papers, and they mentioned these were synonymous with preprints. The lack of consensus in versions of electronic research articles was thus prevalent among studies.

The present study utilizes the definitions of manuscripts and preprints with several modifications. As opposed to manuscripts, the study s employ 'unrefereed' articles, to refer to documents that are yet to be submitted, but not currently under a peer review process. Examples of these documents would be technical reports, working papers, or project reports. Instead of preprints, the present study uses pre-refereed drafts, which refers to documents that are submitted to journals and under review, but not yet accepted. Once the documents are accepted by journals, they are considered 'refereed articles'.

Refereed articles have two versions – author postprints and publisher PDFs. Author postprints are versions closest to publisher PDFs, and usually show some designations, such as 'final version' or 'forthcoming in…'(Antelman, 2006). Swan and Brown (2005) found that refereed, published articles were the most commonly self-archived among research materials. Yet publishers allowing self-archiving rarely permit the author to post publisher PDFs. Interestingly, Antelman reported that around half of 575 self-archived articles sampled from 22 journals in social science disciplines were publisher PDFs. Since publishers' PDFs conveyed more cues about the quality of the articles, and were also more citable than author postprints, authors preferred to use the publishers' versions. Self-archiving behavior, therefore, was affected by norms in disciplines, as opposed to publishers' policies. Another study conducted by Goodman et

al. (2007) compared authors' self-archived manuscripts with published versions to see whether there were significant differences, indicating that self-archived manuscripts might have less validity that the published versions. Based on the comparison of 12 pairs of social science articles and another 12 pairs of Biochemistry articles, the authors found no significant errors that would influence the validity of research from self-archived manuscripts. Thus the authors concluded that authors' self-archived manuscripts had adequate quality, although published versions were more readable and had fewer minor errors than the self-archived versions.

By examining self-archived articles, Antelman (2006) and Goodman et al. (2007) analyzed self-archiving patterns in certain disciplines. The present study, however, investigates individual faculty members as units of analysis and their self-archiving behavior across various disciplines. Additionally, this study examines the range of materials that faculty self-archive including unrefereed articles, pre-refereed drafts, authors' postprints, publishers' PDFs and book chapters, which are more common in humanities (Andrew, 2003).

Motivations of the Study

University faculty members are considered the primary authors of research literature in academia. Faculty expend substantial effort to publish their research work in prestigious publishing venues, implying the quality of the publications, a large readership, and a greater citation rate. Higher citation rates are viewed as signifying greater impact in an area of study and also influence tenure and promotion, which in turn, advances the careers of faculty members.

The existing scholarly publishing mechanisms are complemented by self-archiving behavior as it can improves the communication of research to readers via the Internet. Proponents of Open Access (OA) often mention that this behavior results from

the authors' adoption of OA as an overriding principle – free and unrestricted availability of scholarly content on the Internet. However, it is not known whether or how this principle motivates researchers to self-archive their work. Furthermore, other factors affect and shape self-archiving behavior. For example, Antelman (2006) found most of the self-archived journal articles she sampled to be publisher PDFs. She suggested that the norms of disciplines were more likely to influence patterns of self-archiving than the publishers' policies. Self-archiving practices across disciplines vary by the version and types of research work self-archived and the selection of the multiple forums of publicly accessible websites. Little research, however, characterizes the wide range of faculty members' self-archiving practices, and investigates motivations and barriers to self-archiving. This study attempts to bridge the gap in the literature.

Numerous studies already exist regarding self-archiving behavior in disciplinary repositories (Brown, 2001; Harnad and Brody, 2004; Manuel, 2001; Meyer and Kling, 2002; Lawal, 2002; Pinfield, 2001), as well as research on self-archiving in IRs (Allen, 2005; Foster and Gibbons, 2005; Pelizzari, 2003). Those studies examined self-archiving in one, out of several types of forums, whereas only one study explored multiple self-archiving activities (Swan and Brown, 2005). These studies, however, did not focus specifically on faculty members as the primary subject group. Although most survey respondents of these studies consisted of authors working at universities in various countries, their academic status or rank was not specified.

Furthermore, few studies examine motivating or impeding factors that influence self-archiving behavior. Gadd et al. (2003a) investigated concerns about self-archiving, focusing on copyright issues. According to their study, most academic authors without self-archiving experience worried that their research work may not be published if it had been previously self-archived or that they might violate copyright agreements with publishers. In contrast, authors with self-archiving experience were significantly less concerned about these issues than those with no self-archiving experience. These are

interesting findings, although the study did not cover any other factors that might affect authors' decisions to self-archive. In addition, Swan and Brown (2005) suggested that the majority of their survey respondents were 'self-motivated' to make their research literature publicly accessible on the Internet. Yet, the authors did not make these motivators for self-archiving behavior explicit.

Self-archiving behavior emerged from an environment where (1) conventional publishing systems lack efficiency in delivering research to users in a timely manner; (2) the cost of serials has been soaring making it difficult for many libraries to afford to purchase core journals; and (3) more sophisticated search engine algorithms and metadata harvesting enhance the accessibility of Open Access content. Self-archiving is an innovative approach to disseminating and archiving digital scholarly content. Major funding agencies, such as National Institutes of Health (NIH), encourage or mandate authors to self-archive research papers and associated data sets on designated repositories. University libraries have begun to provide IR services that collect and preserve scholarly output created by members of the university community. In spite of the support and services for self-archiving, faculty participation in this practice is low, especially in IRs. This study addresses this problem by investigating motivations of and barriers to faculty self-archiving behavior.

The study develops the overarching research question – why faculty participation rates are lower than predicted by the literature. Specific research questions are as follows:

- What are existing ways that faculty members make research materials publicly accessible on the Internet?
- Why do they use certain forums for self-archiving?
- What motivates faculty members' self-archiving behavior?
- What makes them reluctant to self-archive their research materials?

In order to answer the research questions, the present study employs two theoretical models: (1) socio-technical network model for scholarly electronic communication; and (2) social exchange theory applied to knowledge sharing behavior. The study investigates existing self-archiving practices, which represent a complex relationship among social and technical factors. The socio-technical network model proposed by Rob Kling focuses on scholarly communication process, which can explain interactions between social factors and technologies involved in self-archiving practices. Social exchange theory has been used as a conceptual foundation to identify motivations of knowledge sharing activities. Self-archiving is considered to be scholars' knowledge sharing and distribution, and thus the theory is appropriate for examining motivators for self-archiving. The following section discusses a conceptual framework centered on those theories.

Conceptual Framework

Self-archiving behavior entails scholarly communication mediated by information technology, particularly digital media and the Internet. Scholarly electronic communication has been explored by an enormous number of studies from various perspectives. Kling et al. (2003) argued that most existing studies could be categorized in one of two models and that neither adequately represented behavior in electronic Scholarly Communication Forums (e-SCFs). The first model was based on two axioms: (1) behavior was affected by the information processing features of technology; (2) actors were considered individual users who choose to, or not to, employ a specific e-SCF. Centered on this model, discussions of e-SCFs were dominated by technological features and efficiencies, and the cultural work practice context of actors was usually ignored. The other model investigated both social and technical aspects, while maintaining clear boundaries between those two aspects. In response to these models, Kling et al. proposed

the socio-technical network model, in order to better explain socio-technical behavior, which referred to "tightly integrated conceptions of the interaction between people and technologies" (p.52). Since self-archiving behavior can be considered a type of socio-technical activity taking place in various e-SCFs, it is appropriate to use the model for the present study.

Although the socio-technical network model is a conceptual foundation of this study, it only addressed the overall aspects of socio-technical networks, as opposed to specific variables which need to be examined. In particular, Kling et al. mentioned the importance of understanding the incentive structures on which the present study concentrates. However, they provided no guidance for investigating incentives. Thus, social exchange theory is applied, since this theory has been utilized in several studies regarding incentives for knowledge sharing in organizations. Hall (2003) in particular, mentioned that social exchange theory was relevant for research on scholarly communication because it represented a social process where actors shared knowledge and had social relationships via research communities. Now that self-archiving is regarded as a knowledge sharing process, social exchange theory is pertinent to determining factors that influence self-archiving behavior. The following sections explain each theory in greater detail.

Socio-Technical Networks

The socio-technical network model characterizes the complex interaction between information processing features of e-SCFs and social behavior. The model presumes that social and technical aspects of e-SCFs are either inseparable or irreducible to one another, but mutually constituted. It was also assumes that participants in e-SCFs are embedded in social relationships, and therefore, have multiple responsibilities. In this respect, the socio-technical network combines participants with different roles, rights,

responsibilities, resource flows, legitimacies, and taboo behaviors (Kling and Callahan, 2003).

The socio-technical approach has been applied to research on digital libraries – viewing them as socio-technical systems – networks of technology, information, people, and practices (Van House et al., 2003). Digital libraries are not just information retrieval systems or digital resources collected on behalf of user communities, but also a component embedded in information-related activities of those communities and several information institutions, such as libraries and archives. The information-related activities include a life cycle of creating, evaluating and seeking documents (Borgman, 2000; 2003). Documents are not merely "things" containing information, but have "social life", reflecting "how, why, when, and for whom people create documents" (Borgman, 2000, p.94). Creating and using documents are also components involved in ongoing evaluation processes conducted by authors, readers, universities, publishers, and libraries. Authors identify publication channels whereas readers decide which documents they will read. Universities make decisions on tenure and promotion based on the quality of faculty members' scholarly work along with teaching and service. Libraries decide what documents they select, collect and provide access for users. This social context needs to be understood in order to develop digital libraries and services. These social issues interacting with technology are dealt with from the perspective of socio-technical networks.

Kling et al. (2003) proposed eight heuristics to model socio-technical networks in e-SCFs: (1) identify system interactors; (2) identify core interactor groups; (3) identify incentive structure; (4) identify excluded actors and undesired interactions; (5) identify existing communication systems; (6) identify resource flows; (7) identify architectural choice points; (8) map socio-technical features to architectural choice points. Based on these constructs, this study focuses on the first six points:

- *System interactors*: The relevant population of actors in self-archiving practices includes academic authors, readers, universities, libraries, and commercial and societal publishers. Since the present study focuses on university faculty's self-archiving behavior, authors encompass faculty members in universities and their co-authors. Readers are any one in the world who has Internet access. Universities are interested in the control of scholarly content as knowledge assets, and specifically they provide economic, technical and administrative support for IRs implemented by university libraries. Publishers entail copyright issues relating to self-archiving pre-refereed draft and published journal articles, or books.

- *Core interactor group*: This study concentrates on perceptions and behavior of faculty members who have performed self-archiving. Methods of self-archiving and factors affecting decisions to self-archive are examined.

- *Excluded actors and undesired interactors*: It is critical to understand the types of interactions that actors do not want to have. This study also explores how faculty members with no experience of self-archiving view such activities and what makes them reluctant to participate.

- *Existing communication systems*: Faculty members can select one or more forums for self-archiving: personal web pages, research group websites, departmental websites, disciplinary repositories, and IRs. The use of existing communication channels for self-archiving may either encourage or discourage the use of a newly introduced system, such as IRs. The present study explores relationships among the use of different forums.

- *Resource flow*: Resource flow has direct or indirect impact on interactions within the network. For example, promotion and tenure committees have direct resource control over untenured professors, and therefore, the views of these committees vis-à-vis self-archiving practices affects whether the untenured professors decide to self-archive their research work. This study investigates how professors perceive the influence of other actors that retain resources and control over their decisions to self-archive.

- *Incentive structures*: This study investigates extrinsic and intrinsic benefits that faculty members perceive when self-archiving their research.

The socio-technical network model provides a general framework that helps us understand the interactions between social and technical elements in self-archiving practices, although it does not mention any specific variables to be examined, especially regarding incentive structures. The incentive issues pertain to factors that motivate or impede self-archiving behavior – the factors on which the present study focuses. In this respect, another theory is necessary to frame those factors in a concrete manner. Social

exchange theory is applicable to research on "motivational factors of knowledge sharing in large, distributed, and information-intensive organizations" (Hall, 2003, p.287). Various studies of scholarly communication, in information science, examined the social exchange of knowledge to some extent, and self-archiving is considered a part of the knowledge sharing process.

Social Exchange Theory

Social exchange theory is congruent with the socio-technical network perspective, because it emphasizes network relations with individuals or groups and exchange of tangible and intangible resources through the relations. Molm (1997) mentioned that social exchange theory arose from a recognition that social interactions existed outside the economic marketplace and that non-monetary social exchanges occurred. These exchanges involve different resources, such as favors between neighbors or votes for political support between politicians. People rely on one another for valued resources and offer them to one another. Social exchanges may take place recurrently based on a history of relations and the mutual contingency of behavior. These recurring exchanges result in patterns of interactions and interdependencies between people over a period of time.

There are key four elements in social exchange theories: (1) actors, (2) resources, (3) structure of exchange, and (4) process of exchange. Actors can be individuals or groups, who behave based on rational decisions to increase positively valued outcomes and decrease negatively valued ones. Resources are defined as the currency of exchange, which would be tangible or intangible. Resources received as a result of the exchange are defined as outcomes. Benefits are positive outcomes resulting from the exchange, whereas costs are negative outcomes in the course of the exchange.

The structure of exchange denotes dependent relationships supporting the exchange. Three types of the exchange structures have also been identified, including

direct and generalized exchange. A direct exchange refers to case in which actor A directly provides value to actor B. In the generalized exchange, actor A and B do not reciprocate directly. Instead, actor B provides value to actor C, who offers value to actor A. Then, actor A provides value to B, although B does not directly return the value to A. Finally, the process of exchange indicates the type of interactions required to conduct an exchange. Followed by 'exchange opportunities' that make the exchange happen, 'exchange transactions' take place, either in a 'negotiated' or 'reciprocal' manner. 'Negotiated transactions' involve reaching an agreement of benefits between actors, whereas 'reciprocal transactions' regard a process in which contributions of actors to the exchange are done separately. Exchange relations, therefore, are developed over time in the reciprocal transactions, when the offerings induced reciprocal benefits (Molm, 1997).

Although social exchange theory does not specify information or knowledge as a type of resource, several other studies have utilized the theory to explain knowledge sharing behavior. Constant et al. (1994) examined factors affecting whether or not people shared their knowledge with a person who had previously behaved derogatively towards them. Based on assumptions in social exchange theory, if people are more self-interested, knowledge sharing is diminished. However, if individuals are predisposed toward a more social attitude and cultural norms for organizational ownership of knowledge exist, information sharing is more likely. Organizational ownership norms were also assumed to be common to people with more work experience. These norms, however, were typical only in a corporate environment, as opposed to an academic setting. The results of these experiments showed that if people believed that their product and expertise were owned by an organization, they were more likely to share them with others who behaved inappropriately.

Jarvenpaa and Staples (2000) tested the factors identified in the study by Constant et al. (1994) - organizational ownership of information and the propensity to share - in a study of academic and administrative staff in a university. The staff shared information

through collaborative electronic media, such as e-mail or list serves. In addition to those two factors - organizational ownership and propensity to share, the authors examined information culture, task interdependence, computer comfort, and perceived characteristics of computer-based information. Based on an analysis of their survey data, all those factors were found to be associated with the use of the electronic media.

In other research, Kankanhalli et al. (2005) and Hall (2001) employed the concepts of costs and benefits in social exchange theory to determine factors affecting contribution to knowledge repositories in corporate environments. Cost factors would result in a reluctance to share knowledge; whereas benefit factors would motivate sharing. Also, both studies examined contextual aspects of knowledge sharing. The categorization of costs, benefits, and contextual factors is relevant to this study, although each factor needs to be interpreted based on self-archiving practices. The following sections identify factors that the present study examines.

Cost and Benefit factors

Kankanhalli et al. (2005) used the social exchange theory to identify factors on an individual level. They focused on resources represented as costs of and benefits from the contribution of knowledge to electronic knowledge repositories. Opportunity costs and actual loss of resources were the main cost issues. In the setting of knowledge repositories, opportunity costs included time and effort to codify and deposit knowledge resources. The actual loss of resources was regarded as a loss of power and unique value within an organization in relation to the transfer of knowledge to the repositories. This actual loss of power, however, would not be relevant to self-archiving practices because authors retain rights and control over their research work even when making it publicly accessible on the Internet. Rather, opportunity costs are appropriate to represent the additional time and effort required for faculty members to input their research/teaching work into forums for self-archiving. Since professors, particularly in research

universities, usually have high workloads, anything that involves additional time and effort can be considered an opportunity cost. Another cost factor relating to self-archiving behavior is uncertainty and concerns about copyright issues (Foster and Gibbon, 2005; Gadd et al., 2003a).

Benefits were identified to be both extrinsic and intrinsic in the study by Kankanhalli et al. (2005). Extrinsic benefits included organizational rewards, reputation, and reciprocal benefits. Organizational rewards indicated increased salary or bonuses, job security, and career advancement. Hall (2001) classified those as explicit rewards for knowledge sharing in organizations. In the study by Kankanhalli et al., organizational rewards were significantly related to knowledge contribution. With respect to self-archiving, however, organizational rewards signify tenure and promotion based on scholarly evaluation, differing enormously from corporate methods of assessment. Thus, the organizational reward factors cannot be directly applied to self-archiving, although it will be interesting to see how faculty members perceive the relationship between self-archiving practices, and tenure and promotion in this dissertation.

Kankanhalli et al. (2005) also proposed that earning a better reputation would be another explicit benefit from knowledge contribution. By reputation, they meant recognition within companies, and it turned out to be unrelated to knowledge contribution. In academic environments, however, reputation in a faculty's own field of study is very important and directly influences tenure and promotion. Therefore, the present study focuses on professional recognition, while institutional recognition is explored, only in the case of faculty members who deposit their research work in IRs.

Reciprocal benefits refer to expected return from knowledge contribution. Reciprocal relationships develop over time, as opposed to a single-time event (Molm, 1997). The mutual interchange of knowledge between authors and readers is hardly expected in self-archiving practices. In fact, self-archiving involves free and unlimited access to research materials, and this open access characteristic of self-archiving implies

a degree of altruism in sharing information. Altruism has been operationalized as satisfaction in helping others through knowledge sharing. This factor was categorized in terms of an intrinsic benefit (Kankanhalli et al., 2005) or a soft reward (Hall, 2001). Faculty members may be altruistically motivated when making their research work publicly accessible on the Internet. However, their self-interest in open access principles would also be a motivator for self-archiving.

In addition to the benefits mentioned in previous studies, I have added three dimensions of scholarly publishing – accessibility, publicity, and trustworthiness - as explicit benefits that might motivate self-archiving behavior. Kling and McKim (1999) suggested that these three aspects were valued by scholars. It will be interesting to see whether faculty members perceive self-archiving practices as supporting accessibility, publicity, and trustworthiness of their research materials.

In sum, the present study applies social exchange theory to self-archiving behavior, and that faculty members have some sense of the costs and benefits they would incur when self-archiving. Based on this assumption, the present study suggests extrinsic and intrinsic benefits relating to self-archiving. Extrinsic benefits include accessibility, publicity, and trustworthiness of self-archived research materials (Kling and McKim, 1999), professional recognition (Swan and Brown, 2005), and academic reward (Kling and Spector, 2003). Intrinsic benefits concern altruistic intention of and self-interest in self-archiving (Cronin, 2005). Cost factors relate to copyright concerns (Gadd et al., 2003a) and additional time and effort required to self-archiving (Foster and Gibbons, 2005). The detailed description of each factor is provided in Chapter 2.

Contextual factors

In addition to the cost and benefit factors, contextual (Kankanhalli et al., 2005) or organizational factors (Hall, 2001) have been proposed as being crucial conditions that motivated or challenged knowledge contribution. Hall (2001) suggested five desirable

conditions for knowledge sharing, which (1) make knowledge sharing as an explicit responsibility; (2) encourage experimentation; (3) value all contributions regardless of their status; (4) promote local communities for sharing knowledge; and (5) provide employees with appropriate technological tools.

Kankanhalli et al., (2005) investigated contextual factors utilizing constructs from an earlier study by Nahapiet and Ghoshal (1998). In order to explain which factors affected the creation and exchange of knowledge in organizations, Nahapiet and Ghoshal concentrated on social capital. They proposed that "networks of relationships constitute a valuable resource for the conduct of social affairs, providing their members with the collectivity-owned capital, a credential which entitles them to credit, in the various senses of the word" (p. 243). Hall (2001) also considered this study as a part of social exchange discourse of knowledge sharing.

The conditions of the knowledge creation and exchange included (1) accessibility of the social knowledge; (2) expectation of values through creating and sharing the knowledge; (3) motivations to create and exchange the knowledge; and (4) combination capability, which referred to abilities to assimilate and use the knowledge. The structural dimensions of social capital were proposed to affect the accessibility and anticipation of values. Nahapiet and Ghoshal suggested that the cognitive dimensions of social capital influenced the anticipation of value and the combination capability. The relational dimensions of social capital were proposed to affect all conditions except for the combination capability. These four conditions needed to be satisfied in order to create new intellectual capital and sustain the dimensions of existing social capital.

A theoretical model representing the relationship between social capital and intellectual capital has been used in other studies regarding contribution to corporate knowledge repositories or electronic bulletin boards of professional associations. In fact, Nahapiet and Ghoshal (1998) emphasized the social-level of knowledge exchange. For example, Nahapiet and Ghoshal suggested the structural dimensions of social capital at

the organizational level, involving network ties and arrangements. Research by Wasko and Faraj (2005) has focused on knowledge contribution at an individual level. Wasko and Faraj suggest that structural social capital can be successfully adapted to the individual level, indicating that an individual's position in the network affects his/her willingness to contribute knowledge to others. Likewise in this dissertation, the relational dimensions of social capital at the organizational level are also investigated in relation to how an individual perceived the effects of relational social capital on his/her knowledge contribution.

Similarly, Kankanhalli et al. (2005) employed the relational dimensions of social capital at an individual level, examining an individual's perception of three relational social capitals including trust, pro-sharing norms, and identification. They also suggested that these three factors were able to augment or reduce the impact of cost and benefit factors upon knowledge contribution. This study, therefore, explores how these three factors of relational social capital, trust, pro-sharing norms, and identification, influence university faculty's self-archiving behavior. Trust indicates belief in good intent and the competence of other actors, such as a university and users. Identification indicates faculty members' concerns with collective outcomes, membership, and loyalty toward universities. However, identification is only relevant to contribution to IRs because the repositories reside in organizational settings, which is not necessarily the case with the other forums for self-archiving. Instead of pro-sharing norms, several articles in the self-archiving literature mention the importance of a pre-print culture, in which researchers distribute drafts of research articles, before they have been peer reviewed, to colleagues around the world. Furthermore, the present study adds one more variable to contextual factors – the influence of other actors on the decision to self-archive, such as co-authors, universities or departments, and grant-awarding bodies. Kling et al. (2003) emphasized this factor in order to understand socio-technical features in e-SCFs. Specific descriptions of each contextual factor is provided in Chapter 2.

In addition to costs, benefits and contextual factors, the present study investigates individual traits as control variables. Few research studies on self-archiving mentioned individual traits. Only Swan and Brown (2005) found that the number of publications was positively associated with participation in self-archiving practices. Other than that, some studies concerning e-journals examined associations of professional rank and administrative positions with faculty members' perception of e-journals (Budd and Connaway, 1997; Lancaster, 1995). Translating these to self-archiving, the present study examines the impact of following three individual traits - the number of publications, professional rank, and administrative positions – upon motivations for self-archiving.

I have consolidated costs, benefits, contextual factors and individual traits in Figure 1, which represents my synthesized model of the factors that influence self-archiving behavior.

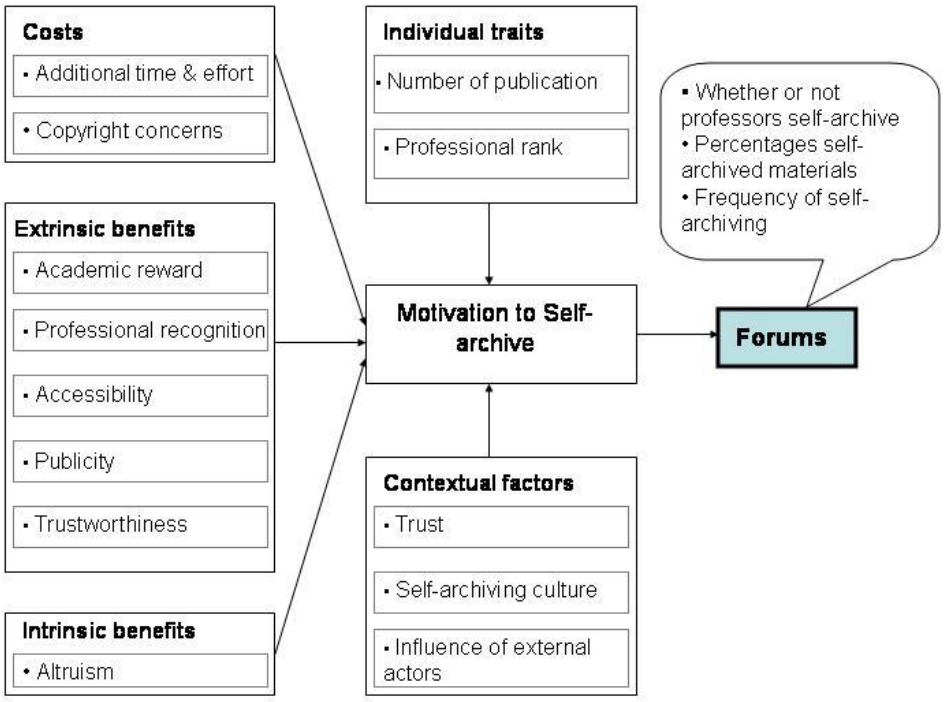

Figure 1.1. Model of Factors affecting Self-archiving Behavior

The model illustrates how various independent variables affect dependent variables, in particular, faculty members' decisions to self-archive. The extent of self-archiving behavior is also measured with respect to the frequency of self-archiving and percentage of self-archived research materials. Cost factors negatively influence the decision to self-archive, whereas benefit factors are assumed to have positive associations with self-archiving. It is also presumed that trust and self-archiving culture as contextual factors, and the number of publications as an individual trait are positively related to the decision to self-archive. The remaining factors are assumed to be associated with faculty members' self-archiving behavior.

Research Design of the Study

The present study employed two methodologies: a large scale survey and then follow up interviews with a smaller group of interviewees. The former method allowed for generalization, the second for more in-depth explanation of phenomena identified in the survey. The survey was conducted first. Based on the survey results, an interview protocol was developed and interviews were performed with 41 survey respondents in order to obtain more in-depth information. Using different methods in one study improves its validity. Interviews demonstrate internal validity for groups of interviewees, although it provides lack of generalizability to the population at large. The problem of external validity is compensated for by the large-scale survey.

The population of participants includes assistant, associate, and full professors of seventeen universities in the U.S. classified as Carnegie Doctorate-granting Universities. Those universities currently have live DSpace IR websites. DSpace is one type of open source software applications most commonly used for IRs. The reason for selecting universities with IRs is that the present study is concerned with all possible methods

available to faculty members for self-archiving, whereas not every university currently provides IR services for faculty members.

Out of the population, two samples are made. One group includes professors whose materials are deposited in their university's IR. Since an IR is a relatively new forum, faculty members are less likely to be aware of IRs and to use them for self-archiving. In this respect, it is interesting to identify early adopters of IRs, and examine their perceptions of and behavior relating to self-archiving. However, it is possible that professors in the sample are not actual contributors to IRs because most IRs in the U.S. are populated with pre-existing research papers collected by librarians (Lynch and Lippincott, 2005). Therefore, actual IR contributors are fewer than the number of professors whose materials are placed in IRs.

The other sample was drawn from the population excluding the IR contributors, based on three prototypical disciplines in each of four areas: science, engineering, social science, and humanities. Within the sampling frame, assistant professors were oversampled. Since the sample of IR contributors was highly skewed toward science and engineering, and tenured professors, sampling professors in other disciplines and/or in tenure-track positions aided the investigation of their self-archiving behavior, and motivations.

A survey instrument for the study was developed, and it included questions regarding the practices that professors use to distribute their research/teaching materials on the Internet, as well as factors affecting self-archiving behavior. The questions investigating factors are presented as statements based on a Likert scale, offering a range of options across five choices: strongly agree, somewhat agree, neutral, somewhat disagree, and strongly disagree.

An interview protocol was developed after the main survey, and validated by interviewing pre-test survey respondents, who agree with being interviewed. The interviews took around 20 minutes by telephone. Semi-structured interviews were

performed to complement survey data. These methods are discussed in greater detail in Chapter 3.

Significance of the Study

The present study has implications for those who are involved in research and practices regarding digital scholarship, digital preservation, knowledge sharing, and information reuse in academic settings. Self-archiving behavior represents a new way of disseminating scholarly content, and understanding this behavior helps extrapolate the transformation of scholarly communication mediated by electronic media and the Internet. Self-archiving behavior can also be viewed as the exchange and reuse of knowledge, although numerous studies of knowledge sharing are concerned with corporate settings. This study contributes to the body of literature as to knowledge management and reuse in academia, particularly in higher education institutions.

The study also benefits self-archiving practices with institutional support – disciplinary repositories and IRs. Through these practices, librarians and archivists have opportunities to expand their roles of organizing and maintaining digital content as a long-term commitment, as well as outreaching other stakeholders in establishing shared goals and policies. In addition, university administrators are concerned about the increasing cost of commercial publications, and therefore, thriving IRs might be a solution to the financial issue. Moreover, publishers who perceive threats against a subscription-based business model may adapt or adopt new models of open access scholarly publishing. The investigation of professors' motivations for self-archiving, and barriers that they perceive, helps all the stakeholders to better design and implement repositories for self-archiving.

CHAPTER 2

LITERATURE REVIEW

Scholarly communication is often used interchangeably with the term scholarly publishing, in spite of the broader conceptions of the former (Kling, 2004). Scholarly communication is a two-way process involving communicators and content. These communicators include authors, readers, editors, scholarly associations, publishers, and librarians. Content ranges from research data and teaching materials to documents that support activities such as journal article refereeing or conference organizing. In addition to the communicative processes, Cronin (2005) shows how scholarly communication has been studied from other perspectives, for example, a functionalist approach to science as a social system (Merton, 1973) or ethnographical research on scientists' work practices in laboratories (Latour and Woolgar, 1979). While scholarly communication is dominated by informal interactions during academic activities, scholarly publishing is "one formal part of scholarly communication" (Kling, 2004, p.593) that scholars ultimately pursue. It may also be a one-way process because numerous research articles are read only by a small number of audiences.

Garvey and Griffith (1972) emphasize the distinction between informal and formal scholarly communication. In their view, formal communication includes various channels of oral communication ranging from small research meetings to large scientific conferences, as well as prepublications, such as technical reports, theses or dissertations, and preprints. These forms of informal communication represent a continuum from very informal, such as journal clubs, to almost formalized, such as conference presentations. In

contrast, formal communication means publishing peer-reviewed research articles in scientific journals, books or reviews, and therefore, it was considered public and orderly. In particular, the orderliness signifies that information in an informal domain must be evaluated via peer review in order to be selected as a quality research article. In addition, citation rates of journals were used to generate impact factors, which represent the quality of journals (Kling, 2004). Based on those quality assurance mechanisms, scholarly publishing comprises a foundation of scholarly evaluation.

These features of scholarly publishing – formal, public, and orderly – are still useful; however, they do not adequately characterize changes in scholarly publishing mechanisms since the early 1990's. According to Kling (2004), early research on scholarly electronic publishing focused on the transition from paper-based to electronic journals (e-journals), and how the use of electronic media benefited scholarly communication. In the late 1990's, most paper-based journals provided electronic counterparts, whereas their business model was based on subscriptions to the print version. Therefore, these print and e-journals (p-e journals) were not viewed as offering the same advantages as pure e-journals, for instance, saving costs, wider audiences, and faster dissemination. More recently, a model of open access journals was proposed by a group of scientists that emphasized free and unrestricted access to journal articles via the Internet (Crow and Goldstein, 2004).

In addition to the variations of e-journals, other scholarly publishing mechanisms or forums have emerged since the early 1990's. The exemplary model is arXiv, an electronic preprint database in high energy physics. In physics, the distribution of paper preprints was an existing practice within an 'invisible college' or narrow circle of closely aligned researchers. The existence of this preprint culture contributed to the rapid adoption of arXiv in physicists' communities. arXiv also fulfilled other information needs of physicists – faster distribution and low cost access. Even though there was no peer review in the system, alternative mechanisms for quality control were suggested, for

example, experts' reviewing e-prints before posting or voluntary readers' comments after posting (Hurd, 1996).

In addition to disciplinary repositories such as arXiv, scholars were posting their research and teaching materials on personal home pages or institutional web pages, although it was often criticized (Cronin, 2005) as a 'vanity press' model due to the lack of quality control. As a result, the term 'self-archiving' was coined to describe this emerging behavior of making research or teaching materials publicly accessible through the Internet. Harnad (2003a) mentioned that self-archiving refereed, published articles was no longer the vanity press model because those documents went through peer review. In other words, authors continued to disseminate their research work via existing scholarly publishing venues, while they posted pre-refereed and refereed articles on publicly accessible websites. This self-archiving behavior now plays a complementary role in scholarly publishing rather than replacing it.

The definition of self-archiving also includes a reference to depositing digital documents in publicly accessible websites, preferably an open access archive or repository. There are two types of open access repositories: disciplinary repositories, such as arXiv, and Institutional Repositories (IR). Several studies also consider personal or departmental websites as another possible forum for self-archiving (Andrew, 2005; Antelman, 2006; Swan and Brown, 2005).

Regardless of the venue, pre-refereed drafts and post-refereed research articles have been found to be the two main types of self-archived digital documents, although various kinds of digital objects have been deposited, particularly in IRs. The forum and the content are two significant components in open access systems. As such, Coleman (2006) argues that it is important to examine where authors deposit their works and which versions of research articles are deposited in order to understand academic authors' self-archiving behavior.

This chapter discusses the aforementioned changes in scholarly publishing incorporated into electronic media and the Internet, focusing on e-journals and self-archiving practices in disciplinary repositories, IRs, and other web spaces. It also considers arguments concerning how the evolution of scholarly publishing has been shaped by both disciplinary cultures as well as co-opted by information technologies (Cronin, 2005; Hurd, 2000; Kling and McKim, 2000; Søndergaard et al., 2003). Interactions between disciplinary differences and information technologies provide a general context for understanding heterogeneous scholarly publishing regimes at this time.

Disciplinary Variations in Scholarly Electronic Publishing

Disciplinary differences in scholarly publishing already existed prior to the widespread use of information technologies, although these differences are often ignored or not emphasized in various studies on scholarly communication mediated by information technologies. According to Kling et al. (2003), such studies focused on the information processing features of technology, as well as on the individual actors who selected to use or not to use a particular communication forum. As a result, discussions of the studies were dominated by features and efficiencies of the forum, without considering the cultural work context of the actors. Kling et al. also noted that studies examining both social and technical components treated these as separate entities, and therefore, interactions between them were not investigated. In response to this gap in the literature, Kling et al. investigated socio-technical behavior representing the interaction between people and technologies. For Kling, this interaction was an essential element for understanding scholarly electronic communication. However, the interaction differed among disciplines because every domain possesses a unique structure including tools,

languages for specific purposes, concepts, meaning, information structure, needs, and relevance criteria (Søndergaard et al., 2003).

In a similar vein, Cronin (2005) suggested "the norms and values that constitute different epistemic cultures shape and. in turn, are shaped by the technologies scholars use to publish their work and to facilitate communication with both their peers and the public at large" (p.194). Knorr Cetina (1999) proposed that epistemic cultures referred to the arrangement and mechanisms in a given discipline to create and warrant knowledge. She investigated "the construction of the machineries of knowledge construction" (p.3), represented in the process of employing a variety of instruments and objects in laboratories, particularly in high energy physics and molecular biology. Knorr Cetina mentioned that this "epistemic machinery" showed different empirical approaches, ontologies of instruments and social mechanisms, and elicited the diversity of epistemic cultures. Cronin explained the cultures represented "quite different socio-cognitive structures, …manifested in differing material practices, communication behaviors, and publishing regimes" (p.3).

Cronin noted that diverse publication practices were shaped by academic reward systems differing in epistemic cultures. For example, molecular biologists valued a number of co-authored short papers in prestigious journals, whereas scholars in philosophy emphasized sole-authored monographs published by a reputable university or commercial press. In addition, quality assurance mechanisms based on peer review vary by discipline. For example, the physicists reading and citing articles in arXiv have confidence in the reliability of those articles. Cronin mentioned that this confidence was rooted on social trust in their professional cultures. However, humanists reside in a more individualistic academic culture, and therefore, they rely on prestigious journals with rigorous peer review procedures, which provided social trustworthiness.

Kling and McKim (2000), in particular, investigated disciplinary differences in the use of e-SCFs, focusing on high energy physics, molecular biology, and information

systems. High energy physicists had disseminated their paper-based preprints since the 1970's, and therefore, this preprint culture enhanced posting electronic preprints on arXiv. Molecular biologists, however, did not have such traditions, and thus, electronic preprint databases did not play an important role in scholarly communication. Instead, they actively used data repositories to share datasets, and these repositories have become critical to their communication. Information systems researchers use another type of web-based collection named ISWORLD, which links to useful resources, course syllabi, and articles and is not considered a repository of full-text articles or data.

The communication practices in scientific fields act as social forces that shape the use of electronic media in scholarly communication (Kling and McKim, 2000). The ways of organizing work and accessing resources to conduct research varies between fields and those differences influence the legitimate forms of communication in such fields. This argument is derived from a perspective called Social Shaping of Technology (SST), which emphasizes the idea that technologies continue to be shaped during their use, and therefore, they are products of human creation and use. The particular shaping of technology occurs in reaction to the needs and values of specific communities. In the realm of science, certain scholarly communities may value the rapid distribution of their published articles to geographically disperse and at the same time, they may be reluctant to make their manuscripts available for the same large community. SST presents a framework that allows for these subtle and highly contextualized preferences for communication to exist. In response to this, Kling and McKim argued that heterogeneity of communication practices among different fields would persist. It was very unlikely that all disciplines would adopt the same ways of using electronic media for scholarly communication. This argument implied that social norms would arise in different fields around electronic publication practices in ways that were not as evident as in the print publication sphere. Hurd (2000) also commented that discipline-specific variations in the adoption of the communication channels were linked to the values of a discipline. The

persistence of these values in a digital environment would lead to the evolution of scholarly communication systems.

Bohin (2004) suggested that the transformation of scholarly communication has not yet been stabilized. The consequence of this change would not be determined by technologies only, but shaped by the response of various actors to new technical solutions. According to Meyer (2006), this Social Construction of Technology (SCOT) framework was consistent with the socio-technical network model that Kling et al. (2003) proposed, in regard to "the identification of relevant social actors, understanding interpretive flexibility, and examining the process of translation and enrollment" (p.38).

Specifically, Bohin examined the case of self-archiving pre-prints in arXiv to demonstrate interpretive flexibility of the repository. He pointed out that the perceived benefit of a new technology relied on "cultural patterns and social conventions in the communities in which it is introduced" (p. 379). He postulated that the perceived value of pre-print repositories, such as arXiv, varied in disciplines, depending on (1) whether pre-print culture existed; and (2) variation in journal acceptance rates. This second criterion was important to explain why scholars in certain disciplines self-archived pre-prints. It was known that journal rejection rates were the highest in Humanities, the lowest in the Physics and in between in Social Sciences. Bohin postulated that scholars in disciplines where journal acceptance rates were low, considered pre-prints to be unrefined manuscripts and thus they felt constrained referring to pre-prints. However, researchers in disciplines where papers were unlikely to be rejected, believed that there was no substantial difference between preprints and published versions, and accordingly they were willing to share preprints. Bohin assumed that the diverging journal acceptance rates would influence self-archiving not only pre-prints, but also post-prints, i.e. published articles.

The following sections investigate the "plurality and plasticity" of scholarly publishing practices exhibited in electronic journals and self-archiving behavior.

Electronic journals are the primary publishing channels for scholars, whereas self-archiving is comprised of supplementary activities, which precede or follow journal publication. Publishing journal articles and self-archiving those articles reinforce, rather than compete with, each other. The next section discusses definitions of electronic journals and factors affecting submission rates. After that, studies regarding forums for self-archiving, especially disciplinary repositories and IRs are examined. Finally, factors affecting self-archiving behavior are illustrated.

Electronic Journals

Definitions

According to Kling and Callahan (2003), in the early 1990's electronic journals (e-journals) were based on the following assumptions: (1) they would be pure-electronic; (2) they would be free to their authors and readers; (3) they could be peer-reviewed. Relying on these assumptions, concerns about long-term preservation of e-journals and their trustworthiness were raised. These assumptions, however, are no longer valid since most publishers currently provide electronic versions of paper journals, rather than relying on solely electronic media. Also, the journals remain to be distributed based on a subscription model, which allows individuals or organizations to access electronic copies of journal articles by purchasing licenses. This allows electronic journals to capitalize on their paper-based counterparts that already have a reputation and readership. Distributing electronic versions of the journals does not damage their legitimacy. For instance, electronic versions of highly prestigious journals, such as *Nature*, do not provide any evidence of losing their legitimacy due to its electronic distribution. The legitimacy issue of e-journals primarily concerns only pure e-journals.

Kling and Callahan claim that the distinction between e-journals without a paper version and a paper journal with an electronic version is critical when explaining such

issues as the legitimacy of e-journals or their costs. However, most studies regarding e-journals do not differentiate varying formats of e-journals and therefore, the sundry types conflate into one category, 'e-journals'. Kling and McKim (1999) noted that this conflation causes misunderstandings surrounding the changes in the use of e-journals. For example, some empirical studies reported that an increasingly number of researchers used e-journals (Harter, 1998; Rusch-Feja and Siebeky, 1999). The e-journals, in fact, referred to electronic versions of paper-based journals distributed by conventional journal publishers as opposed to pure e-journals. Kling and Callahan, as a result, suggest four categories of e-journals as follows:

- Pure e-journals: journals that are disseminated only in digital form
- E-p journals: journals that are primarily distributed in digital form, but which also produce limited paper versions
- P-e journals: journals that are primarily distributed in paper form but which disseminate electronic versions
- P+e journals: journals that began with parallel paper and electronic versions

Recently, Open Access (OA) journals, which make their content freely available on the Internet, have emerged. Willinsky (2006) mentioned that OA journals were considered a breakthrough approach, which allowed for wide access to research work with value and quality, although this method implies various legal, political and economic issues unresolved. They are not based on subscriptions, but on different business models. Swan and Brown (2004) mention that a primary model for OA journals makes authors pay an article-processing fee, and then the publishers make their articles freely available online. The Public Library of Science (PLoS), a non-profit corporation for OA journal publishing, implements the author-fee model. Another OA journal publisher, BioMed Central (BMC) operates both the author-fee model, and an institutional membership model. This makes authors' institutions join BMC for a fee so that authors can publish articles in BMC journals without a payment. As a consequence,

the JISC (Joint Information Systems Committee) pays for membership for all UK higher education institutions.

However, in some cases established journals based on a subscription model have converted to OA journals. McVeigh (2004) reports that more and more journals in the ISI citation databases are adopting the OA distribution model. Thus, OA journals are not necessarily new, although most established journals using OA distribution make only a few recent years of content freely accessible on the Internet. According to Crow and Goldstein (2004), those planning to convert to OA journals should consider whether their publishing medium is electronic-only, or both electronic and print. This decision involves cultural and economic issues. Depending on the disciplines the journals serve, authors, readers or markets may resist electronic-only journals. In this case, maintaining dual media for journals becomes the better strategy. If the journals previously offer only paper-based volumes, they need costs to implement an electronic publishing capability. If they have already used both publication media, costs depend on whether they go for electronic-only or both. If OA journals are not electronic-only, but an electronic and paper-based hybrid, they can be classified in all four categories – pure e-journals, e-p, p-e, or p+e journals - proposed by Kling and Callahan. In this respect, the categorizations are not relevant to OA journals, and therefore, OA journals should be considered another type of e-journals. In the following sections, therefore, factors affecting the submission to e-journals are examined, separating e-journals from OA journals. After that, factors influencing the adoption of OA journals are discussed.

Factors affecting E-journal Submission

Academic Reward System

One of the major forces that motivates scholars to publish their research work is a reward system, which provides peer recognition and tenure or promotion for scholars as

authors. The reward system is closely related to scholarly journals, which play an important role in distributing the results of research quickly and widely, and in preserving them over a long period of time. The increasing visibility of the scholarly work through journals helps colleagues recognize new ideas of the research. In addition, scholarly journals perform a quality control function through a peer-review process, and therefore, an author's publications in scholarly journals demonstrate his/her research performance, as well as the quality of the publications. The wide adoption of traditional scholarly journals in research communities has been strongly influenced by the fact that scholarly journals serve as a standard instrument for academic reward systems (Kim, 2001)

In the case of e-journals, however, researchers have reluctance to publish their papers there because the academic reward system in many universities is not as open to e-journals as to traditional scholarly journals. Thus scholars are unwilling to publish their research in e-journals until universities or research communities demonstrate that they will reward publications in electronic journals (Budd and Connaway, 1997; Kim, 2001; Lancaster, 1995). Budd and Connaway (1997) found that only 13.6% of assistant professors perceived an openness in their universities to accept publications in e-journals for tenure and promotion purposes, whereas 30.3% of associate professors and 37.7% of professors did perceive some degree of acceptance of e-journals. These findings indicated low incentives for faculty to adopt e-journals to publish their research.

In addition, Sweeney (2000) discovered that 67% of the faculty members he surveyed agreed that electronically published articles should be counted in the tenure and promotion process. Some of the faculty members who agreed with the statement mentioned that they felt a burden to demonstrate the quality of the e-journals in which they published their research. In order to do so, they documented the acceptance rate and the review process of the e-journals. Other faculty who also supported the idea of counting e-journals in the tenure and promotion process suggested that if an e-journal had been widely adopted as a quality publication, it would not matter whether it was

published in paper or electronic form. These findings indicated that the quality and prestige of e-journals, which are directly associated with academic reward systems, were main concerns in their adoption. The quality issues are examined in the following section.

Prestige/Quality of E-journals

The prestige and quality of scholarly journals were identified as main factors that affect researchers' journal selection in reference to submission of their papers, as well as the reading of journal articles (Gomes and Meadows, 1998; Kim, 2001; Schauder, 1994; Speier et al., 1999). Gomes and Meadows (1998) examined what scientists perceived to be important characteristics of e-journals and p-journals. The authors analyzed 213 questionnaires completed by physicists, chemists, and biologists. 82% of respondents reported journal prestige as p-journals' most important characteristic; while 64% of the respondents reported content quality as being the chief importance. However, only 0.8% and 0.4% of respondents reported that journal prestige and content quality, respectively, were the most critical elements of e-journals. 57% of respondents reported that publication speed was a crucial attribute of e-journals. Those findings showed the perceived lack of prestige of e-journals, which made scientists reluctant to submit their research articles to e-journals.

Similarly, Speier et al. (1999) investigated the quality of e-journals perceived by faculty members in several business schools by using self-administered questionnaires. Questions using a 7-point Likert scale were used to measure their perceptions. The results showed that faculty members perceived electronic peer-reviewed journals to be of lesser quality than paper peer-reviewed journals. Even electronic versions of paper journals were not perceived to be as high quality as their paper counterparts. This perception showed that respondents' belief about e-journals was based on pure e-journals or e-p journals (Kling and Callahan, 2003).

Usability of E-journals

Bishop (1995) examined early forms of seven e-journals from a reader's point of view. The journals were pure-electronic and freely accessible via ftp, listserv, Gopher or browser and therefore, they were different from current formats of e-journals. She argued that the critical problems with e-journals were their usability, lack of instructions, and low readability. Readability was also found to be important for users of e-journals in another study. Steward (1996) mentioned that at least 84% of respondents in her study wanted e-journals to provide adequate type fonts, text design, and layout, as well as capabilities of browsing text/graphics, and of underlining and annotating. Liew et al. (2000) also suggested that the difficulty of reading on screen and inability to highlight or make notes were factors that hindered the use of e-journals.

In addition to readability, searching and browsing capabilities were important for scholars (Brennan et al., 2002; Rowland et al., 1997). Brennan et al. (2002) noted that faculty members used e-journals to find articles for which they had a citation. They also searched for articles by author or subject within a specific title or group of titles. Hyperlinks embedded in articles enabled faculty members to navigate full-text articles in various journals, and they considered this functionality very desirable. According to Mayernik (2007), internal and external hyperlinks to citations were commonly utilized at 11 pure-e journals in Psychology and 10 in Physics. However, the use of multimedia was relatively low as a feature of pure e-journals - only 120 instances of multimedia were found and 93% were videos.

Factors affecting OA Journal Submission

Several studies investigated factors that impact authors' submission to OA journals (Björk, 2004; Rowlands et al., 2004; Park and Qin, 2007; Swan and Brown, 2004; Warlick and Vaughan, 2007). In particular, Rowlands et al., and Swan and Brown

conducted international surveys of authors to examine their awareness, motivations and concerns about OA journals. Both studies identified the lack of awareness of OA journals as a primary reason why authors did not submit their articles to these journals. Rowland et al. (2004) reported that out of 3,787 respondents across 97 countries, 34% knew nothing about OA journals and 48% had a little knowledge. In contrast, Swan and Brown (2004) stated that out of 157 non-OA journal authors, 62% were aware of the concept of OA journals. However, the main reason that non-OA journal authors did not publish in OA journals was that they were not familiar enough with OA journals in their fields. This finding points to the need for effective promotion of their existence, although Björk (2004) mentioned that most OA journals lacked resources for marketing, partly due insufficient understanding of marketing tactics.

Both the study by Warlick and Vaughan (2007) and the research by Park and Qin (2007) were based on the analysis of interview data, although the latter was more theory-driven research than the former. Warlick and Vaughan interviewed fourteen professors in Biomedical Science at UNC-Chapel Hill and Duke who published in OA journals between January 2004 and June 2005. The interview data were collected in-person, via e-mail or telephone. The authors found that the faculty measured publication quality according to the Impact Factor or the prestige of journal. These were considered to be the most important factors even when deciding on an OA publication venue.

Likewise, Park and Qin (2007) suggested that perceived journal reputation was one of the factors that influenced authors' willingness to publish in OA journals. Park and Qin analyzed interviews from eight faculty members and six doctoral students at Syracuse University in using a grounded theory approach. Their findings suggested that three factors – (1) perceived journal reputation, (2) perceived topic relevance, and (3) perceived availability of OA journals – affected interviewees' willingness to publish and use articles in OA journals. The authors also mentioned that previous studies regarding cognitive authority and topical relevance identified those three factors as related to the

usage of publications. Still, these "well-established" relevance criteria were associated with both publishing and using OA journals.

Motivations

Swan and Brown noted that the main reason that OA journal authors publish their articles in the journals was that they supported OA principles – free and unrestricted accessibility of their articles to the public. This option of 'free to access' was found to be the most appealing concept associated with OA journals (Rowlands et al., 2004). Swan and Brown also stated that OA journal authors were motivated by the that the fact OA journals often publish articles faster, have large readerships, and consequently, greater citation rates to their articles. Similarly, Warlick and Vaughan (2007) suggested that the incentives to publish in OA journals were free and easy access by diverse audiences and broad dissemination of the online articles. Yet, the authors noted that free access and increased visibility were not strong enough motivations to choose OA journals over subscription-based journals unless their publication quality was ensured.

Interestingly, non-OA journal authors in the study by Swan and Brown had an opposite perception toward OA journals – smaller readership and a lower citation rate. In addition, the second main reason not to publish in OA journals was their low prestige and low impact. This result was somewhat contradictory to the finding that 49% and 46% of OA journal authors believed that high prestige and high impact, respectively, were important reasons to publish in OA journals. However, Warlick and Vaughan suggested that even OA journal authors had mentioned that OA journals had a lower impact and this made them reluctant to publish in OA journals.

Concerns

Quality and preservation of content in OA journals were found to be two primary concerns (Rowlands et al, 2004; Swan and Brown, 2005). On the quality side, new OA

journals need time to develop a decent impact factor and most OA journals with impact factors were not ranked highly. McVeigh (2004) examined impact factors of 239 OA journals covered by ISI citation databases. Fourteen journals ranked in the top 10% based on their impact factors, while two-thirds of the journals were below the 50th percentile in rank. The quality of journals was also related to peer review. Swan and Brown reported that 98% of OA journal authors and 95% of non-OA authors, believed peer review was an important feature of scholarly journals. Authors wanted the peer review process in OA journals to be as rigorous as that of traditional journals. Similarly, Suber (2002) suggested that the quality of journals was dependent on the quality of the editors, referees and authors, and the high standard for those participants should be applied to OA journals. Swan and Brown noted that few authors agreed with the author-fee model, which discourages OA journal submission by authors from developing countries or those having no grants to cover the fee.

Preservation is another issue. Swan and Brown mentioned that 42% of non-OA journal authors were worried about the preservation of OA journal content. On the other hand, Rowlands et al. stated that their survey respondents disagreed with the idea that OA journals were ephemeral. Although the concern about preservation "should not be a good reason for eschewing OA journals" (Swan and Brown, 2004, p.67), preservation of OA contents was an important issue and needs to be investigated more in depth.

If OA journals do not step up to accomplish preservation activities, these may be left to libraries or individuals. Thus, the OA practice termed self-archiving draws the attention of researchers and practitioners engaged in scholarly communication. Academic authors self-archive scholarly content in publicly accessible websites including personal or departmental web spaces, and disciplinary or institutional repositories. Following sections examine studies regarding those forums, and factors affecting self-archiving behavior as well.

Self-archiving Practices

Disciplinary Repositories in Science

Disciplinary repositories exist and are actively used in scientific fields, such as physics, mathematics, and chemistry. A successful example of the repositories is the arXiv, which was initiated at the Los Alamos National Laboratory in 1991 and currently migrated to the Cornell University Library. arXiv is a free electronic manuscript archive, a key resource for physicists and its goal is to "turn the electronic media into a means of author empowerment" (Taubes, 1996, p.767). Most of the studies investigating disciplinary e-scripts repositories focus on authors' use and participation in this exemplary archive (Brown, 2001; Harnad and Brody, 2004; Lawal, 2002; Manuel, 2001; Meyer and Kling, 2002; Pinfield, 2001). In particular, Meyer and Kling (2002) focused on the leveling effect of the arXiv in fields participating in the repository from the perspective of the socio-technical network model. The majority of the studies investigating arXiv used bibliometric methodology.

Brown (2001) investigated citation patterns of e-prints posted in arXiv and the extent to which journal editors allowed citation to e-prints, as well as journal instructions regarding the citation. Physics is a field that has adopted e-prints much more widely than other disciplines. Brown mentioned that the low adoption of e-prints in other fields might be due to the lack of peer-review in the e-print repositories. With respect to citation patterns, she used citation data from SciSearch and examined citations to e-prints of 12 repositories of arXiv from 1991 to 1999. 35,928 citations were made to the arXiv e-prints for a citation rate of 34.1%. She also investigated citations to the e-prints between 1998 and 1999 in 37 physics and astronomy journals and found substantial overlapping in citations between disciplines of physics and astronomy, and therefore, citations were not limited to the archive designated to the same field as journals. Journal editors' survey responses indicated that although they accepted the citation of e-prints, their attitude was

not enthusiastic. Journal instructions to authors suggested that citation of e-prints tended to be dependent on the policies established by journal publishers. Thus, policies of citing e-prints and subsequent publications vary from journal to journal. Brown concluded that the citation of e-prints showed some possibility that peer-review would be limited or even eliminated in physics and astronomy.

While Brown (2001) focused on the citation pattern of e-prints in arXiv, Harnad and Brody (2004) investigated the impact of those Open Access (OA) articles in arXiv, comparing with non-OA journal articles. The study, therefore, compared citation counts of individual OA and non-OA articles, as opposed to examining impact factors of journals. Harnad and Brody obtained 14 million articles in physics from 1992 to 2001 covered by ISI citation databases. Among those articles, they found 260,000 articles deposited in arXiv. The percentage of OA articles gradually increased from 1% in 1992 to 18% in 2001. Citation counts of OA articles were greater than those of non-OA articles during the 10-year period, as the ratios of citation counts of OA articles to those of non-OA articles were between 2.5 and 5.7. This finding indicated larger impact of OA articles than non-OA articles.

Another bibliometric study was conducted to investigate various aspects surrounding e-prints including types and country of authors, level of collaboration, citation of other e-prints, level of publication in traditional peer-reviewed journals, and eventual transformation of e-prints into refereed publications (Manuel, 2001). Manuel selected three e-print repositories for the study including MPRESS, arXiv, and SLAC/SPIRES HEP. MPRESS included e-prints in mathematics, whereas SLAC/SPIRES HEP stored e-prints from high-energy physics. She randomly sampled 100 e-prints in each archive and examined aforementioned aspects. Most authors of the sampled e-prints were from academic departments. Mathematics featured a low percentage of co-authors, whereas high energy physics demonstrated a high percentage. International collaboration in authorship was observed in around 25% of the sampled e-

prints. E-print authors also actively published their research in traditional peer-reviewed journals. For example, authors of e-prints from arXiv published 5.38 research articles on average in peer-reviewed journals. Out of the sample e-prints from arXiv, 50% were published in peer-reviewed journals. This result is contrary to a common assumption that almost all e-prints were published in traditional journals. Based on this assumption, Björk (2004) suggested that legal issues and academic reward issues were not significant factors in e-print repositories.

Manuel (2001) also found that the majority of authors in those repositories were from developed countries and this indicates that creating and disseminating e-prints is not as equitable, or as an efficient a model for distributing research output to scientists worldwide as expected,. She mentioned that this contradicted the hypothesis that e-print repositories would increase the participation of scientists in peripheral positions, such as in developing countries or less research-intensive universities. Meyer and Kling (2002) challenged the "leveling hypothesis" by examining posting rates and institutional affiliations of authors on sampled e-prints in arXiv. 1,329 e-prints posted in two years - 1993 and 1999 - were randomly sampled from three repositories of arXiv, including HEP-TH (High Energy Physics-Theory), MATH and Astrophysics. The authors were categorized based on the academic status of their institutional affiliation as classified by the Carnegie Foundation. In 1993 and 1999, 79.9% and 88.6% of authors that contributed to the HEP-TH were from Research I universities, which were the most elite universities in the Carnegie classification. Research I universities awarded at least 50 doctoral degrees a year and received more than 40 million dollars in federal support every year. In the MATH repository, 74% in 1993 and 72% in 1999 were from Research I universities. In the Astrophysics repository, 45.9% in 1993 and 62.6% in 1999 were from other countries, most of which were developed countries, such as Germany, France and Japan. Thus, there was low participation of scientists from less research-intensive universities or developing countries.

Meyer and Kling (2002), therefore, suggested that the leveling hypothesis was founded on the perspective that considered only two roles of scientists, authors, and readers and their interactions with arXiv. From the perspective of the socio-technical network model, however, each scientist was a complex actor with various social roles, such as a teacher, a collaborator, a member of a scholarly society or an employee of an academic institution. Understanding the relationship of arXiv with these other roles was necessary to identify the likelihood of scientists' contributions to arXiv, as well as factors external to such behavior.

Pinfield (2001) examined the use of arXiv and some implications for the implementation of institutional repositories. arXiv accepted non-proprietary formats because physicists worked on UNIX, and therefore they did not often use word-processing software. Self-submission was also used in arXiv and e-mail alert service of newly posted e-prints was offered. Informal peer-reviews in arXiv were also valuable for authors to revise their e-prints for acceptable publications. Most of the publishers in physics accommodated arXiv e-prints, and therefore, they adapted their copyright agreements to permit self-archiving. For example, the Institute of Physics required authors to sign over copyright but granted 'personal license', which enabled authors to post e-prints into non-publisher servers. Pinfield suggested that physicists used arXiv to distribute both e-prints and post-scripts, and at the same time, they wanted to publish their research work in traditional peer-review journals. Therefore, arXiv was used as a supplementary publishing system for many scholars.

Lawal (2002) investigated reasons of use and non-use of disciplinary repositories by using questionnaires. She sampled 473 researchers in various disciplines including chemistry, biological science, engineering, cognitive science and psychology, mathematics and computer science, physics and astronomy. Survey results indicated that 51.6% of responding physicists/astronomers and 28.8% of responding mathematicians/computer scientists used e-print repositories. However, over 90% of

respondents in engineering, cognitive science/psychology, biological science and chemistry did not use e-print repositories. In particular, all of the chemists were self-reported non-users. Main reasons of non-use of e-print repositories were found to be publishers' policies and technical constraints. Primary reasons for use of e-print repositories were rapid dissemination of information and increasing visibility of their research work. This reason was consistently found in another study that investigated a disciplinary repository in chemistry. Brown (2003) examined the usage and acceptance of the Chemistry Preprint Server (CPS) and the majority of responded users of CPS mentioned wide and rapid distribution as a main reason for the usage. However, only 6% of responded editors of chemistry journals accepted e-prints in their journal publications, and this low acceptance of e-prints by journals resulted in no citations of e-prints in the traditional journal literature.

It should be noted that the majority of literature that I reviewed in this section concerned arXiv. The preponderance of focus on arXiv may not be generalizable to other disciplinary repositories.

Working Paper Series: Guild Publishing Model

Few studies exist regarding working paper series created in academic departments. These valuable scholarly assets exist throughout academia and represent a publishing model for scholarly work currently implemented in various research institutions. Kling et al. (2002) suggested that the Guild Publishing Model (GPM) was based on the practice of academic department or research organizations, which published series of working papers, technical reports, research memoranda and occasional papers on the Internet. This publishing model was most common in artificial intelligence, computer science, mathematics, economics, linguistics and physics. The authors analyzed three conceptions of GPM: (1) business models, (2) localized site and access, and (3)

quality indicators. GPM websites are freely accessible by readers, as well as authors sponsored by individual departments and institutions. Therefore, the costs of operating GPM are seen as affordable for many academic departments and institutions, although the costs depend on existing infrastructure, such as technical access and support.

Since the GPM was based on needs and resources of individual institutions, GPM websites are locally controlled, rather than limited by rigid rules or standard formats. Access to the GPM websites is possible through searches by authors or institutions. However, if the websites are not indexed by search engines, locating the websites is not possible. In addition, GPM performs career review, which was a mechanism for quality control that differed from peer-review. The quality of e-prints in GPM, therefore, relies on the reputation of sponsoring institutions or on the entry criteria of membership in the institutions. For example, Harvard Business School (HBS) research manuscript series allows papers authored or co-authored by the HBS faculty. Since the entry barrier of HBS was high, the legitimacy of the research manuscripts was considered high.

Several studies mentioned departmental websites as a forum where authors commonly self-archive their research materials (Andrew, 2003; Hey, 2004; Swan and Brown, 2005). These studies however did not specify how these forums were operated or their characteristics. The departmental websites may or may not represent the GPM that Kling et al. (2002) suggested.

Disciplinary repositories and working paper series produced in academic departments share several issues with IRs, such as quality control or legal concerns. For example, the career review system in working paper repositories is similar to the quality control mechanism for IRs. Journal publishers' policies of posting their manuscripts or preprints affect authors' decisions to contribute to those forums. Authors' acceptance of IRs, however, is more complex than that of disciplinary repositories, because IRs are constructed via the participation of authors, in various disciplines of a university, whose needs and social practices differ.

Institutional Repositories

An Institutional Repository (IR) is another place for self-archiving, which provides Open Access (OA) to the scholarly digital content created by a university community, typically represented as departments, institutes, centers and laboratories. In addition, IRs aim to focus on "the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment" (Lynch, 2003, p.2).

Shearer (2003) suggested that the success of IRs would eventually be determined by "their uptake and use by researchers" (p.106). She pointed to the critical mass of content that led to the significant usage of disciplinary e-print repositories. Translating this to IRs, she argued that the success of an IR should be determined by its use, and one of the measures for the usefulness of IRs was contribution of content. While faculty members are regarded as the important contributors to IRs, several studies have found a low rate of faculty submission to IRs (Chan, 2004; Foster and Gibbons, 2005; Pelizzari, 2005; Young, 2002).

Chan (2004) conducted a case study of TSpace, an IR of the University of Toronto. During a one year implementation period, Chan realized that the submission rate for the faculty's scholarly work was low. This has also been observed in other universities. She suggested cultural inertia — the faculty's resistance to change in the means of disseminating their research work – might affect the low submission rate. She also mentioned lack of awareness of the importance of open access and lack of trust in the institutional commitment to long-term access. Moreover, Chan (2004) and others (Björk, 2004; Crow 2002) have identified uncertainty about intellectual property and copyright restrictions as concerns of faculty members.

Several empirical studies have confirmed the aforementioned low faculty participation in the IRs (Allen, 2005; Davis and Connolly, 2007; Foster and Gibbons,

2005; Xia, 2007). Foster and Gibbons (2005) conducted a qualitative study that investigated the work practices of faculty members at the University of Rochester, in order to identify what the faculty members wanted in an institutional repository. The authors videotaped faculty work practices in situ, conducted interviews, and invited faculty to show how they identified, used and distributed scholarly work in digital form. The study participants were 25 faculty members from five disciplines including economics, physics, political science, linguistics, and visual and cultural studies. Their findings suggested that the primary need of faculty for use of an institutional repository was that other people were able to find, use, and cite the work they submitted to the repository. If these criteria were not satisfied, even enthusiastic supporters of the institutional repository would not use them. Foster and Gibbons also identified other reasons why faculty did not rush to submit their content. One reason of non-use was closely related to the faculty work practice, such as co-authoring or versioning. Each faculty member developed their own routine to create and organize documents in progress, although the institutional repository based on DSpace supported only the capture of finished work. Some faculty members also mentioned that the name of 'institutional repository' implied that the system supported the needs and goals of the institution, as opposed to those of individual users. Due to these points, faculty lack incentives to contribute their research work to IRs.

Davis and Connolly (2007) performed a case study that evaluated the DSpace IR at Cornell University, particularly focusing on the reasons for faculty non-participation. The authors interviewed eleven faculty members across various disciplines and examined their attitudes and behaviors relating to self-archiving in disciplinary or institutional repositories. In fact, nine out of the eleven professors used personal or research group Web pages for self-archiving research work. The primary reasons for using the Web pages were ease and control. Some departments provided Web pages with the same design for all faculty members to promote the department and its faculty to prospective

students or to the public. Faculty members who have not self-archived in any OA repositories stated several reasons for not using such venues: (1) no perceived benefit from using a new system, (2) copyright concerns, (3) materials in a repository had variable and questionable quality, (4) journals might prohibit the submission of research work deposited in a repository, (5) fear of plagiarism and getting scooped, and (6) reluctance of self-archiving pre-refereed or unrefereed research work, which would not be as accurate as peer-reviewed work.

Only four interviewees were aware of the Cornell DSpace IR and only one of them had contributed to the repository. Reasons for the non-participation included (1) disciplinary repositories already met the needs of faculty, (2) lack of DSpace functionality, for example, users could not delete, move, or cross-list items across categories, (3) the IR was not associated with a narrow research community, and (4) IRs were perceived to be isolated resources so that users searched materials in an individual IR separately. The fourth reason was based on the misconception of the IR's search capabilities since metadata for the IR's materials were indexed and could be searched by Internet search engines. The authors suggested that faculty members' perceptions about the risks and benefits of self-archiving in repositories were formed by "disciplinary norms and their reward structure". Addressing this cultural diversity across disciplines would be necessary in order to encourage faculty contribution to the IR.

Another empirical study regarding IRs was conducted to investigate humanities scholars' awareness and attitude toward IRs (Allen, 2005). Similar to the present study, Allen used a survey and follow-up interviews. He distributed a web survey questionnaire to 5,500 individuals who subscribed to humanities sections in JISC listservs. Only 75 responded to the survey. Around 79% of these respondents were from universities, half lecturers or professors, and the rest were PhD or post-doctoral students. Over 60% of the respondents worked at institutions in the U.K. Forty-nine of the 75 had contributed to their universities' IRs, while the remaining respondents had not. Almost two-thirds were

found to be IR contributors, and this might be due to the fact that most respondents were from U.K. institutions where IRs were more widely adopted than other countries and there are policies that require or encourage submission. A majority of both contributors and non-contributors perceived increased accessibility to be an advantage of IRs, whereas the potential for plagiarism and breaking copyright agreements were perceived as disadvantages. It was impossible to generalize these findings to all humanities scholars due to the low response rate. However, this study is the only existing research on humanities' perception on IRs, despite its problems with external validity.

The three empirical studies above suggest a wide range of reasons for the use and the non-use of IRs. Xia (2007), however, assessed faculty self-archiving in IRs by examining the rate of deposits from faculty members in four disciplines – Physics, Economics, Chemistry, and Sociology – across seven IRs: three in Australia and four in the U.K. The author focused on testing the assumption that researchers in disciplines where subject-based repositories exist, such as Physics and Economics, were more likely to make contribution to IRs. The depositing patterns discovered in the study, however, showed no evidence that supported this assumption. This finding was consistent with that in the study by Davis and Connolly (2007). These authors noted that the use of disciplinary repositories made faculty consider IRs redundant, and consequently faculty tended not to self-archive in IRs.

In addition, Xia asserted that two operational strategies of IRs were factors that increased faculty participation. One was "a liaison system" (p.648) or a proxy submission, which freed faculty from the extra time and effort required to deposit materials into IR. Watson (2007) similarly noted that out of 21 authors she interviewed, 16 mentioned the value of "mediated deposit". Xia examined a metadata element that indicated who actually deposited a certain material in the IRs. He found that IRs with an effective liaison system, which assigned one or two administrative staff for inputs, resulted in a high rate of deposit from faculty members. The author suggested that the

liaison practice played an important role in enlarging the amount of content in IRs. The second strategy was the implementation of a mandate policy. Among the seven IRs that the author examined, the Queensland University of Technology (QUT) implemented this policy and it generated a positive outcome – the highest percentage of deposits from faculty among the IRs studied. Xia suggested that a mandate policy was needed for the success of IRs.

However, Kennan and Wilson (2006) pointed out that "while the greatest participation and growth in content comes from those institutions which have a mandatory deposit policy, it is acknowledged that mandating alone is probably not successful" (p.243). The authors quoted an e-mail message from the IR manager at QUT, which indicated that other than the mandate policy, the IR team provided several services for faculty, for example, checking publishers' policies on self-archiving to address the uncertainty of copyright issues, and allowing faculty to upload files in any format. Kennan and Wilson emphasized that social and behavioral factors other than the mandate policy influenced faculty contribution to IRs and such factors needed to be explored.

These studies suggest factors affecting the contribution to each of these forums - disciplinary repositories, working paper series, and IRs. Although some factors are only relevant to specific forums, most variables identified can be integrated under the umbrella of self-archiving behavior and apply to all venues for self-archiving. When examined en masse, these factors can be categorized as follows: (1) cost factors; (2) extrinsic benefits; (3) intrinsic benefits; (4) contextual factors; (5) individual traits.

Factors affecting Self-archiving Behavior

Each group of factors – costs, extrinsic benefits, intrinsic benefits, contextual factors and individual traits – contains one or more variables identified based on the

literature regarding self-archiving, knowledge sharing, and scholarly publishing. The following sections discuss each variable in detail.

Cost Factors

Copyright concerns

Several studies concerning self-archiving suggested that uncertainty about copyright is one of the barriers that impedes the participation in self-archiving practices (Allen, 2005; Chan, 2004; Gadd et al., 2003a; Foster and Gibbons, 2005; Shearer, 2003). In particular, Foster and Gibbons noted that the faculty members they interviewed did not want to violate any copyright issues in the course of their IR contributions.

Copyright issues in self-archiving practices were thoroughly investigated in the RoMEO (Rights Metadata for Open-archiving) project. Gadd et al.(2003b) conducted a survey that examined academic authors' ownership of copyright. 542 academic authors from 57 countries responded. Half of the respondents were scientists, 38% from the social sciences and humanities, and 12% from engineering. One-third of the respondents did not know who owned the copyright for their research papers; whereas 61% thought they owned the copyright. Gadd et al. noted that this finding was not surprising since other studies regarding copyright ownership provided similar results – indicating that most academics rarely consider copyright issues. Academic authors were even less aware of their institutions' copyright policies. In addition, 50% of the respondents stated that 71-100% of their research papers were multi-authored. This finding has been interpreted to mean that the greater the number of authors, the more likely disagreement about self-archiving would occur. Furthermore, 25% of respondents mentioned that they had cleared third-party content in order to publish a paper. This result suggests that for a fairly substantial number of authors, copyright questions are complicated and that authors

would need to ask permission from third-parties, not only to publish a paper in learned journals, but also to self-archive it.

Additionally 49% of respondents reluctantly assigned their copyright to publishers, whereas 41% assigned copyright to publishers freely (Gadd et al., 2003b). This contradicts some earlier data cited about authors' not knowing who owns the copyright. Gadd et al. continue by arguing that the large proportion of authors assigning copyright to publishers indicates that academic authors value the publishers' responsibilities to protect their work from misappropriation by others and to act against any copyright infringements. Also, publications increase the possibility of getting tenure and promotion, which leads to an increase in income. Gadd et al., however, mentioned that authors were not aware that the copyright assignment for the protection of rights also allows publishers to have exclusive economic rights to handle their published work.

Hoorn and van der Graaf (2006) also distinguished between moral rights and exploitation rights when discussing copyright issues in OA journal publishing. They attested that moral rights were generally well-acknowledged in scholarly communication, whereas exploitation rights involved issues of use and reuse of scholarly materials if they became Open Access. The redistribution of an original article relied on publishers' permission since most publishers retained exclusive copyrights. Other than this traditional model of copyright, Hoorn and van der Graaf identified three emerging models implemented in OA journals: (1) a model in which authors retain the copyright in full, (2) a model in which authors share the copyright with the Creative Commons license and this license provides broad permission for the reuse of an article, and (3) a model in which authors transfer exploitation rights that only enables publishers to distribute articles first and to claim all commercial rights later. Authors retained all other rights. Based on the analysis of Web surveys from 335 OA journal authors, Hoorn and van der Graaf found that around 50% of survey respondents preferred the first model, which supported authors' retention of copyright. Copyright retention, therefore, is a factor that

encourages authors to adopt OA publishing, although Warlick and Vaughan (2007) found that only three out of 14 OA journal authors that they interviewed perceived copyright retention to be an incentive to publish in an OA journal.

Additional time and effort

Foster and Gibbons (2005) mentioned that faculty members tended to "resist clerical responsibilities, and resent any additional activity that cuts into their research and writing time". They also did not want to do anything complicated when contributing to the IR. Similarly, Van House (2003) noted that researchers she interviewed at the University of California-Berkeley (UCB) were concerned about the additional work necessary in cleaning data sets and creating documentation and metadata prior to contributing these data to the Digital Library (UCB-DL). The additional work was called 'productizing work', which faculty perceived as requiring new skills different from creating 'real' data. Productizing work was necessary, according to the researchers, to prevent potential criticism about their research work.

In the context of IRs, entering metadata is considered a time-consuming task, which discourages authors' contributions to an IR. Carr and Harnad (2005) addressed the issue of extra effort by measuring the amount of time taken to submit items to the repository in the School of Engineering and Computer Science at the University of Southampton. They analyzed three months of log files for 260 new e-prints' deposit sessions. The average time for metadata entry was found to be 10 minutes 40 seconds, demonstrating that deposit to the IR was not time-consuming. The authors asserted that the investment of this small amount of time would result in a desirable outcome, greater interoperability of Open Access repositories. However, it is uncertain whether academic authors actually perceive 10 minutes per article not to be time consuming and adding 6 articles equals one hour. The authors did not determine the average number of articles entered into the repository, so the average amount of time may be larger than 10 minutes.

Extra time and effort to enter metadata is difficult to be justified unless authors clearly see the benefits from Open Access repositories. In addition, Greenberg et al. (2002) mentioned that 11 participants in their experiments perceived entering metadata to be fairly easy and to have positive attitude toward the value of adding metadata, while this finding cannot be generalized.

Extrinsic Benefits

Accessibility

Kling and McKim (1999) suggested three dimensions of scholarly publishing as communicative practice – accessibility, publicity, and trustworthiness. These were used to assess how effectively an article or a book had been published in a scholarly community. In this sense, self-archived research materials can also be evaluated centered on these dimensions.

Accessibility of a document is defined as any that enables readers "to access the document independent of the author, and in a stable manner, over time" (Kling and McKim, 1999, p.897). For long-term and stable accessibility, institutionalized stewardship was necessary and it was enhanced by shared standards, indexing mechanisms and professional practices. While Kling and McKim criticized the lack of stewardship in electronic publishing practices, IRs are an exceptional case emphasizing organizational stewardship of digital documents created by members of a university community. DSpace IRs, in particular, provide persistent identifiers to all deposited documents, and thus each document has a unique and unbreakable URL. In addition, the Open Archives Metadata Harvesting Protocol (OAI-PMH) is a shared standard used in IRs and some disciplinary repositories, such as arXiv or DLIST. The OAI-PMH was a mechanism for harvesting XML-based metadata from repositories, and therefore, it makes possible interoperable search and retrieval among repositories (Branin, 2005).

Moreover, the accessibility of an Open Access article via search engines was an important factor that would encourage its use. Nicholas et al. (2007) investigated the use and users of a journal, *Nucleic Acids Research* (NAR) in the period of moving from a subscription model to OA, based on the analysis of the usage logs. The authors suggested that although the usage of *NAR* articles increased after the journal became OA, the most significant growth of the use was seen when the journal website was "opened up fully to search engine indexing" (p.876). This finding suggested that the OA impact was moderated by how users searched for OA articles.

Other than IRs and disciplinary repositories, however, most forums used for self-archiving are not based on various technologies that improve long-term accessibility of digital content. In my analysis I discuss two variables related to the selection of a venue for self-archivist. First, whether faculty recognize the advantages of IRs or other OAI-compliant repositories and their concern about long-term preservation and secure maintenance of their research work self-archived on the Internet.

Publicity

Publicity of research literature is defined as having primary and secondary audiences acknowledge the availability of the document. It represents a series of activities from subscription, report lists, abstract databases, advertising and special issues, and citations (Kling and McKim, 1999). Kling and McKim cast doubt on the effectiveness of the publicity in electronic publishing practices, such as self-archiving. Compared to prestigious journals with well-established readership and reputation, electronic publication has thus far had much less publicity to increase the awareness of researchers in a scholarly community.

However, several recent studies suggested greater citation rates for Open Access (OA) articles than non-OA articles (Antelman, 2004; Harnad and Brody, 2004; Lawrence, 2001). Lawrence (2001) analyzed citation rates of 119,924 conference articles

in computer science, and found that on average, freely available articles had a 2.6 times greater number of citations than articles in off-line journals. Antelman (2004) also examined the number of citations in OA articles and non-OA articles in four disciplines – philosophy, political science, electrical and electronic engineering, and mathematics. She reported greater mean citation rates of OA articles than non-OA articles and the difference in citation rates was statistically significant. Specifically, Harnad and Brody (2004) employed references from 14 million articles in 7,000 journals in the ISI databases from 1991 to 2001, as well as from 260,000 articles in arXiv. The authors compared the number of citations of OA articles and that of non-OA articles. The ratio of the citations of OA articles to those of non-OA articles ranged from 2.8 to 5.1, which indicated greater impact of OA articles than non-OA articles. Those studies all indicated that OA articles had more citation counts than non-OA articles, and therefore, OA articles had more impact. Based on this, several studies suggested that materials in self-archiving forums would provide potentially high impact on scholarship (Crow, 2002; Jones et al., 2006; Swan and Brown, 2005).

However, Kurtz et al. (2005) suggested that there were at least three hypotheses that could explain the greater citation rate of OA articles in arXiv than non-OA articles: (1) the open access hypothesis that free and unrestricted access to OA articles enables authors to read and cite them more frequently, (2) the Early View (EV) hypothesis that postulates that since an OA article appears earlier, it obtains primacy and additional time of exposure to users, which generates more citations, and (3) the self-selection Bios (SB) hypothesis which suggests that authors tend to post "the most important, and thus most citable articles" (p.1396) publicly available on the Internet. By analyzing citation data from the NASA Astrophysics Data System and arXiv, Kurtz et al. identified strong EV and SB effects, but no OA effect in Astronomy. This result disputes previous studies that found an OA impact. These three hypotheses have also been examined in other studies. Moed (2007) noted that there were SB and EV impacts, but no OA impact existed, by

analyzing citation rates of arXiv preprints in Condensed Matter Physics. However, Davis and Fromerth (2007) explored citation rates of arXiv preprints in Mathematics and found no OA and EV effects but some inferential support for SB effects. Interestingly, all three studies failed to identify any strong OA effect. This signifies that self-archiving a research article by itself does not increase its citation rate. However, there is some evidence that other factors – early dissemination of an OA article and its quality – do appear to influence citation rates.

Trustworthiness

In the context of self-archiving, trustworthiness of a document concerns the social processes that ensure the quality based on community-specific norms. In academia, trustworthiness has traditionally been determined by peer review. Studies have suggested that academics have concerns about self-archived research work with respect to the absence of a quality control mechanism, which has been identified as the most important value-added service provided by peer review in scholarly journals (Pelizzari, 2005; Swan and Brown, 2005). In particular, another study investigating academic value systems associated with scholarly publishing and communication, highlighted that faculty members considered peer review as "the hallmark of quality that results from external and independent valuation" and "an effective mean for winnowing papers" that faculty members read in the course of their research (King et al., 2006, p.4). King et al. claimed that peer review was a deeply-embedded value in the scholarly publishing system so that any new scholarly publishing model without peer review would eventually fail to be accepted by faculty.

However, Cronin (2005) attested that peer review was viewed as a continuum ranging from double-blind peer-review to open peer review. Research materials could also be evaluated based on the prestige of journals as well as authors' reputations. Additionally, their affiliating institutions have likewise been indicators of

trustworthiness. These criteria may be employed by faculty members when they judge the quality of self-archived research work.

Van House (2002) investigated the credibility of biodiversity data set in a publicly accessible digital library, CalFlora. She argued that "networking allows information to cross social and technical boundaries that have previously provided the context for assessing its credibility and meaning, throwing into relief practices of trust. Sharing information requires that users and providers trust one another. Networked information raises questions at each end of the information transaction" (Van House 2000, p.100). Users of CalFlora needed to assess credibility of information largely depending on assessments of individuals who contributed to their work. Three criteria were found to be used: (1) competence based on methods and content of the work, (2) honesty and lack of deception, and (3) shared orientation and values. Van House suggested that understanding membership in communities and epistemic cultures of science implied similarities of practices and shared meaning, which helped identify individuals who can be trusted.

Academic reward

Academic reward systems include tenure and promotion which is based on research performance and other factors, such as teaching and service. Quantity and quality of publications are major indicators of this performance, although impact of one's research is increasingly an important factor on tenure and promotion. Cronin (2005) cited numerous studies that indicated a strong correlation between citation and research performance rankings, such as RAE (Research Assessment Exercises) in the U.K., as well as associations between citation and faculty salary. In spite of the evidence of citations as proxies for research performance, Tenure and Promotion Committees in the U.S. and U.K. relied on citation data much less than peer review procedures of academic departments.

Similarly, Kling and Spector (2003) suggested that citation counts were considered one of the indicators of the quality and impact of publications. In addition, the place of publication, such as peer-reviewed journals, and book reviews were used to evaluate scholarly publications for tenure and promotion purposes. Those indicators, however, were "well known to be imperfect" (p.94), despite being employed widely as effective criteria. Kling and Spector argued that publications of electronic versions of manuscripts (e-scripts) on the Internet should be assessed carefully based on whether or not those e-scripts were peer-reviewed, in order to be counted toward tenure and promotion.

In fact, a study regarding the University of California faculty attitudes and behaviors regarding scholarly communication suggested that faculty members "overwhelmingly rely on traditional forms of publishing, such as peer-reviewed journals or monographs" (UCOSC, 2007, p.4), which the current tenure and promotion system values. The study pointed out that the academic reward system impeded faculty adoption of a new publishing model. It was also found that once faculty members published an article or a monograph, they tended to be less concerned about disseminating their final publications. This indicated that faculty considered "the act of publishing itself to be sufficient for accomplishing their goal" (p.5). Since self-archiving is related to the process of dissemination after publishing, an academic reward structure that emphasizes conventional publishing is a factor that impedes the faculty participation in self-archiving practices. The present study will investigate faculty's perceived effect of self-archiving behavior on his/her tenure and promotion.

Recognition

According to Ziman (1984), professional recognition in science takes three major forms: (1) publications in reputable scientific journals, (2) citations of research work by other scientists, and (3) attribution of ideas and concepts to researchers. Due to the

problem of information overload, it is very difficult for scientists to receive attention from other researchers, thus making professional recognition "ever-strong currency" in science (Cronin, 2005, p.7). Although trustworthiness and prestige in self-archiving forums are questionable, some studies suggested that increased visibility of researchers' work could be a major benefit provided by self-archiving practices (Crow, 2002; Swan and Brown, 2005). Crow attested that IRs enabled broader dissemination and increased use of contributed materials. This, in turn, drove professional visibility and awareness. Swan and Brown (2005) also mentioned that academic authors' primary objectives for publication were to communicate their research findings to peers, and therefore, peers were able to build upon the results. Professional recognition, therefore, was closely related to the publicity factor regarding citation and potential impact of research work.

Faculty members are not only members of their scientific communities, but also members of a university community. Thus, institutional recognition is also of interest to university faculty. This construct applies primarily to IRs. Knowledge management studies have suggested that by contributing to knowledge repositories, employees expected enhanced reputation in their organization (Hall, 2001; Kankanhalli et al., 2005). However, those studies are concerned with a corporate environment different from a university atmosphere where faculty members are "free agents" rather than "workers for hire" (Branin, 2005, p.5). Thus, faculty members are assumed to be more interested in professional than institutional recognition. Depending on the tenure system and the number of tenure and promotion reviewers who are outside one's discipline, there might be more of an incentive for institutional recognition in some universities.

Intrinsic Benefit

Altruism

Harnad (2006) claimed that "there is an element of golden-rule reciprocal altruism underlying self-archiving, insofar as user-access alone is concerned, but when it comes to author-impact, self-archiving is a matter of pure self-interest" (p.9). In a similar vein, Tschider (2006) noted that OA publishing was regarded as "the act of gifting" scientific knowledge, which in return, resulted in citations of OA articles. In this respect, OA publishing "can be both self-interested and somewhat altruistic". Cronin (2005) also suggested that self-interest and altruism connected to the increasing adoption of OA publishing and self-archiving, because academic authors were interested in making their articles easily accessible. Cronin stated that authors' self-interests in OA publishing and archiving reflected their needs for "branding, competition, and vanity" (p.33) in scholarship. In addition, Odlyzko (2002) mentioned that whether academic authors liked it or not, they were involved with a "war of the eyeballs" (p.18) to which the ease of access was critical. Concerning the idea of reciprocity, Wasko and Faraj (2000) suggested that when people considered knowledge as a public good, they expected "generalized reciprocity", defined as "help given to one person is reciprocated by someone else, not by the original recipient of help" (p.169). They also saw knowledge sharing as motivated by moral obligation, rather than by self-interest.

Kankanhalli et al. (2005) also suggested that knowledge contributors might be motivated by satisfaction with their altruistic behavior, that is, helping other people by depositing content to knowledge repositories. Merton (1988), however, argued that "free and open communication" in science was an institutionalized practice, as opposed to altruism. By "free and open" communication, Merton meant traditional publication systems in which scientists obtained recognition by peers, the essential extrinsic reward in science. Yet, Cronin noted that he was "inclined to think that Merton would have welcomed the communicative transparency of open-access publishing" (p.7), since Merton suggested an idealized structure of science based on four norms - universalism, communism, disinterestedness, and organized skepticism. Among these, communism was

closely related to OA publishing and archiving, because it was based on the idea that scientific knowledge was commonly owned. Originators of ideas could not utter how or by whom those ideas were to be used, although they were able to obtain credit for their effort and creativity. Thus, some argue that the results of research should be publicly accessible (Sismondo, 2004).

However, Merton's normative structure of science has been criticized by other researchers. One criticism is that the actual science does not seem to be governed by Merton's norms. For example, a scientific community in a highly competitive area might have a norm of secrecy, the opposite of communism. Secrecy may even be valued in such an environment because it allows researchers to concentrate on research without worrying about other scholars doing the same study. The other criticism regards a flexible interpretation of Merton's norms, which results in the lack of analytical power of the norms. Others claim that Merton's norms could be used to justify any scientific actions. Sismondo, therefore, suggested that norms were no longer viewed as constraining actions, but as rhetorical resources. Based on these arguments, the approach to altruism in the present study focuses on individual intentions, as opposed to group norms.

Contextual Factors

Trust

Van House (2002) mentioned that trust had been investigated in various areas, such as philosophy, sociology and political science. Although she focused on epistemological trust in networked information – credibility and evaluative criteria for web resources, there were other approaches to examine trust. One approach emphasizes calculative trust, which concerns assessments of risks anticipation of others' behavior, and rational choice. The other regards the role of trust in the social order relating to citizenship, cooperation, and morality. By embracing these two approaches, Nahapiet and

Ghoshal (1998) defined trust as "a willingness to be vulnerable to another party" (p.254). This willingness resulted from four beliefs: (1) the good intentions and concern of exchange partners, (2) their competence and capability, (3) their reliability, and (4) their perceived openness.

Some studies have suggested that authors might have distrust of users' good intentions and competence in using self-archived materials appropriately (Crow, 2002; Davis and Connolly, 2007; Polydoratou, 2008; Van House, 2003; Van Raan, 1997). Van House (2003) stated that CalFlora's data providers were particularly concerned about the misuse of data, due to (1) users' lack of competence required to reuse data appropriately and (2) users' utilization of these data to refute the data owners' original research. In particular, Polydoratou (2008) noted that theoretical chemists that she interviewed were reluctant to share data before completing their research, owing to the possibility of infringement of their work and misinterpretation of data. All interviewees, however, were willing to make their data publicly accessible after finishing their research projects. Van Raan (1997) also noted that scientists would never post their pre-prints or other types of digital information including their premature ideas on the Internet, because it was possible for someone to take over their ideas without proper citation. Similarly, Crow (2002) and Davis and Connolly (2007) mentioned that the possibility of plagiarism and criticism of non-peer reviewed work made contributors reluctant to deposit their work in IRs.

Specifically in the case of IRs, faculty contributors need to have trust in their institutions and the integrity, wisdom and competence of people who manage and preserve work submitted to IRs (Lynch, 2003). In this respect, Chan (2004) noted that if faculty lacked of trust in the library's long-term commitment to IRs, they would be less likely to contribute. Faculty also worried that they would forfeit ownership of their work by depositing it to IRs (Hess and Ostrom, 2004).

Self-archiving Culture

Swan and Brown (2005) described researchers in pre-print cultures as distributing drafts of research articles before they have been peer reviewed to colleagues around the world. The purpose of making their pre-prints publicly available was summarized as the three following aspects: (1) to establish ownership of the research, (2) to develop a certain area of study, and (3) to request commentary prior to final revision and submission of the articles to scholarly journals. As previously noted, Kling et al. (2002) examined another type of pre-print practices called the 'Guild Publishing Model', where working papers, technical reports, research memoranda, and occasional papers are the norm. This publishing model is most common in artificial intelligence, computer science, mathematics, economics, linguistics, and physics. Bohin (2004) also suggested that the culture of sharing pre-prints was constructed, dependent on journal acceptance rates varying in disciplines. Several other studies have hypothesized the positive relationship between pre-print culture and the adoption of OA venues, although the present study examines self-archiving of not only pre-prints but also other types of research materials, especially published articles. Therefore, this study will examine the relationship between the existence of 'self-archiving culture' – sharing research materials openly on the Internet – and the participation in self-archiving practices.

Influence of Other Actors

Faculty members' decisions to self-archive research work can be affected by other actors who have power over resources, such as money or skills. Kling et al. (2003) suggested that identifying who controlled valuable resources in a socio-technical network helped understand the reasons that actors participated in a certain forum for scholarly communication. Swan and Brown (2005) found that respondents to their survey were unlikely to be influenced by co-authors, authors' institutions, and grant-awarding bodies when they decided to submit their work to OA journals. However, a few open-ended

responses from the survey indicated that institutions and grant-awarding bodies did not accept OA materials when funding was considered. Some institutions only regarded ISI-listed journals as 'true' research literature. According to Davis and Connolly (2007), however, the few professors that they interviewed mentioned that their grant-awarding bodies, such as NIH, strongly encouraged them to deposit manuscripts and related data sets. These findings indicate that positive attitudes of universities or grant funders toward self-archiving would provide an incentive for faculty members making a decision to self-archive.

Individual Traits

Individual characteristics that might relate to self-archiving include faculty ranks, and the number of publications. Tenure-track professors have more pressure on them to publish their research work through prestigious channels, and therefore, self-archiving practices would not be their priority. They might also be less motivated to self-archive in the IR of their current university than tenured professors, because it is possible for them to leave for another institution when tenure is not granted. It is assumed that, therefore, perception of tenure-track professors on self-archiving practices differs from that of tenured professors. In addition, the number of publications would affect authors' decisions to self-archive. Swan and Brown (2005) suggested that the more publications authors made per year, the more likely it was that they self-archive their research materials. The investigation of these individual traits will provide implications to design policies and services of IRs and disciplinary repositories.

In conclusion, eleven factors and two individual traits, which might affect self-archiving behavior were identified based on literature review (Table 2.1). The conceptual relationship between the aforementioned factors and self-archiving behavior are empirically determined based on data collected from a survey and interviews. Since these

factors cannot be directly observed, faculty's perception of these factors is measured by developing scale questions. In addition to employing survey methodology, I designed an interview protocol to collect qualitative data that reflected faculty's opinions and experiences of self-archiving practices. The precise methodology is described in the following chapter.

Table 2.1. Factors affecting self-archiving identified from literature

Cost factors	*Benefit factors*	*Contextual factors*	*Individual traits*
Copyright concerns	Accessibility	Trust	Number of publication
Additional time and effort	Publicity	Self-archiving culture	Rank
	Trustworthiness	Influence of external actors	
	Professional recognition		
	Academic reward		
	Altruism		

CHAPTER 3

METHODOLOGY

This chapter presents how the theoretical model presented in Chapters 1 has been broken down in concepts and furthermore how survey and interview questions were developed from these concepts to test hypotheses. As stated in Chapter 1, research questions of this study are as follows:

(1) What are existing ways that faculty members make research materials publicly accessible on the Internet?

(2) Why do they use certain forums for self-archiving?

(3) What motivates faculty members' self-archiving behavior?

(4) What makes them reluctant to self-archive their research materials?

In order to look for answers to these questions, specifically exploring motivations for self-archiving, I developed a series of hypotheses concerning the relationship between independent variables and a dependent variable – faculty members' decision to self-archive. Table 3.1 shows those hypotheses.

Table 3.1. Independent Variables and Corresponding Hypotheses

Factors / Independent variables		Hypotheses
Costs	Copyright concerns	Authors' copyright concerns are negatively related with the decision to self-archive.
	Additional time and effort	Additional time and effort are negatively related to the decision to self-archive.
Extrinsic benefits	Accessibility	Accessibility is positively related with the decision to self-archive.
	Publicity	Publicity is positively related to the decision to self-archive.
	Trustworthiness	Trustworthiness is positively related with the decision to self-archive.

	Academic reward	Academic reward is positively related with the decision to self-archive.
	Professional recognition	Professional recognition is positively related with the decision to self-archive.
Intrinsic benefits	Altruism	Altruism is positively related to the decision to self-archive.
Contextual factors	Trust	Trust is positively related to the decision to self-archive.
	Self-archiving culture	Self-archiving culture is positively related to the decision to self-archive.
	Influence of external actors	Influence of external actors is related to the decision to self-archive.

Using a combination of quantitative and qualitative methods, I tested these hypotheses in examining motivations for and barriers to self-archiving. The order of this chapter is: (1) target population and sampling strategies; (2) survey instrument development; (3) survey data gathering; and (4) interview data gathering.

Target Population and Sampling Strategies

The population of participants includes assistant, associate, and full professors of seventeen universities classified as Carnegie Doctorate-granting Universities, which offer at least 20 doctoral degrees per year. The 17 universities were identified in September 2005, from the list of live web sites of DSpace IRs, provided by the DSpace federation. DSpace is an open source software application widely adopted by university IRs in the U.S. Since the present study concerns a wide range of self-archiving practices by faculty members, it is desirable that professors being surveyed would have as many available options for self-archiving as possible. Not all universities provide IR services at present, and in this respect, the population is limited to faculty members in those universities having DSpace IRs.

The list offered by the DSpace federation presented 110 institutional repositories across 22 countries. Thirty-four repositories were found to be implemented within the universities in the U.S. Among those 34 universities, eighteen universities were Carnegie Doctorate-granting Universities – sixteen were categorized as Research Universities-Very

High research activities (RU-VH), whereas two were classified as Research Universities-High research activities (RU-H). However, I excluded one from the 18 universities because I planned to conduct a pilot survey in the university. As a result, the population included assistant, associate and full professors at the 17 universities.

The total number of the population was determined by asking reference librarians in each university for the total number of assistant, associate and full professors. Counts were available for 12 of these universities. For those universities, I used the total number of faculty publicly displayed in the university web sites. The total resulting population equaled 28,287.

The present study samples two groups of faculty members: (1) faculty members whose materials are deposited in their universities' IRs and (2) faculty members whose materials are not deposited in IRs. In fact, this distinction between contributing or not contributing to IRs may not be meaningful at this time, because the initial collection development of IRs in the U.S. was usually accomplished by libraries' strategy of ingesting existing bodies of technical reports or other gray literature available on campus (Lynch and Lippincott, 2005). This implies that even though faculty members' materials are deposited into IRs, those faculty members may not have had anything to do with the actual deposit and therefore, do not realize those materials are in IRs. Nevertheless, it is still possible to determine actual contributors to IRs among those professors whose materials are placed in IRs. Understanding the early adopters' perceptions toward and behavior of self-archiving would be important, as it helps inform the investigation of self-archiving behavior across various forums.

In order to identify faculty contributors to IRs, I searched author names displayed in IRs via online directories of the 17 universities, and saw whether or not the authors were assistant, associate, or full professors. In addition, if several professors co-authored a given item, I considered each of them to be an individual contributor. The resulting IR contributors included 621 faculty members.

Faculty non-contributors to IRs are included in the whole population, apart from the faculty contributors. These faculty members may self-archive their research materials using other methods. The sampling frame of these faculty members was focused on three prototypical disciplines in each of four areas: science, engineering, social science and humanities. Some of the disciplines are known to have pre-print culture. For instance, physics, mathematics, and biology were selected as disciplines in science. It is common to distribute pre-prints in physics and mathematics, whereas it is not in biology. In addition, sociology, psychology, and economics were selected as social science disciplines. Pre-print repositories are widely used in economics, while they are uncommon in sociology and psychology. For engineering, mechanical, electronic engineering, and computer science are selected. A large number of faculty contributors to IRs are affiliated in those disciplines. In humanities, English literature, history, and art are selected. Very few humanities' scholars were identified as IR contributors, and therefore, examining perceptions of humanities scholars helped understand how professors in those disciplines view self-archiving practices.

The sampling frame was established by collecting the list of faculty members from web sites of those 12 departments in each university. After developing the sampling frame, a disproportionately high number of assistant professors was sampled. The primary task of assistant professors is to achieve tenure based on the quantity and quality of publications. This tenure pressure influences the opinion of assistant professors toward self-archiving, which may differ from that of tenured professors. Since the IR contributor group includes few assistant professors, oversampling assistant professors from the sampling frame resulted in a sufficient number of those to explore their perception and behavior. As a result, the total number of IR non-contributors was 879. By addition the non-IR contributors to 621 IR contributors, the total sample contained 1,500 faculty members.

Survey Instrument Development

Survey questions were developed based on the literature regarding self-archiving practices. Specific questions were adapted from surveys used to examine self-archiving behavior (Allen, 2005; Gadd et al., 2003a; Ober, 2005; Rowlands et al., 2004; Swan and Brown, 2004, 2005) and contribution to knowledge repositories (Kankanhalli et al., 2005).

The questionnaire consists of four sections regarding (1) self-archiving methods, including two yes/no questions about self-archiving experience and awareness of IRs in their universities, 13 multi-choice questions, and one open-ended question; (2) perception of self-archiving, containing 36 Likert-scale questions and one open-ended question; (3) plans to self-archive in the future, including one yes/no question about whether to self-archive in the future, one multi-choice question, eight Likert-scale items about reasons for future contribution to IRs, and one open-ended question; and (4) a demographic section that consists of 11 questions. The survey instrument is presented in Appendix A. The questionnaire is created in a web survey mode.

Measurement of Dependent Variables

The extent of self-archiving experience was measured through several dimensions: (1) whether or not faculty members ever self-archived their research/teaching materials; (2) frequency of self-archiving in the last 3 years; (3) length of time that they have been involved in self-archiving; (4) percentage of self-archived research materials in the past 5 years. Among these possible dependent variables, I utilized percentages of self-archived research work as a measurement of self-archiving behavior. This variable represented what proportion of research materials that faculty members created have been publicly accessible on the Web. The types of research materials included (1) pre-refereed articles; (2) refereed articles; (3) unrefereed articles;

(4) book chapters; (5) data sets. The percentages of these materials that faculty made openly accessible indicated the degree to which professors have participated in self-archiving practices.

The other dependent variable employed in this study involved whether or not faculty members have contributed to their university IRs. The survey instrument provided a question regarding the frequency of self-archiving in several Web venues including IRs. Based on responses to this question, I identified real IR contributors, in order to compare their perception of self-archiving with that of others who have not deposited into IRs.

Measurement of Independent Variables

This study is concerned with faculty members' perception of self-archiving, which cannot be observed directly. In order to measure people's subjective states, scales are commonly used. Scales refer to "collections of items combined into a composite score…to reveal levels of theoretical variables not readily observable by direct means" (DeVellis, 2003, p.8-9). Fowler (1995) mentions that multi-item scales generate more detailed and reliable measures of a common underlying variable than a single question by pooling information that items have in common. Given that the present study attempts to measure various constructs categorized by costs, benefits, and contextual factors, using multi-item scales is appropriate. Table 3.2 presents operationalized definitions of each factor that this study will assess.

Table 3.2. Operationalized Definitions of Costs, Benefits and Contextual Factors

Factors / Independent variables		Definitions	Sources
Costs	Copyright concerns	The extent to which professors perceive requirements to ask permission from publishers, possibilities of copyright infringement, and frictions among co-authors in self-archiving research work	Gadd et al., 2003a, 2003b

	Additional time and effort	The extent to which self-archiving is perceived to require time and effort	Foster and Gibbons, 2005; Van House, 2003
Extrinsic benefits	Accessibility	The extent to which professors perceive self-archived materials to be available in a stable manner, over time	Kling and McKim, 1999
	Publicity	The extent of perceived readership and citation rate of self-archived materials	Kling and McKim, 1999
	Trustworthiness	The extent to which professors perceive self-archived materials to have credibility	Cronin, 2005; Kling and McKim, 1999
	Academic reward	The extent to which professors perceive self-archiving to influence tenure and promotion	Cronin, 2005; Kling and Spector, 2003
	Professional recognition	The extent to which professors perceive self-archiving research work to increase visibility in their field	Cronin, 2005; Swan and Brown, 2005
Intrinsic benefits	Altruism	The extent to which professors have altruistic motivation to self-archive	Cronin, 2005; Kankanhalli et al., 2005
Contextual factors	Trust	The extent to which professors perceive readers' good intentions and competence in using self-archived materials	Van House, 2002; Van Raan, 1997
	Self-archiving culture	The extent to which professors perceive that it is common to self-archive, in both their field and their academic department	Kling et al., 2002; Swan and Brown, 2005
	Influence of external actors	The extent to which professors perceive the influence of co-authors, grant-awarding bodies and universities on the decision to self-archive	Kling et al., 2003

Each factor was measured by using two to four items presented as statements, followed by 5-point Likert scales that range from 'strongly agree' to 'strongly disagree'. Additionally, the option of 'I don't know' is provided for certain questions where respondents may not have enough knowledge to express a belief statement. Items corresponding to each factor are presented in Table 3.3.

Table 3.3. Concept Map

Factors/ Independent variables		Survey Items
Cost factors	Copyright concerns	• I need to ask permission from publishers to post my work on publicly accessible web sites. (item1) • If I post my work on publicly accessible web sites without permission, I may infringe on copyright. (item2) • I need permission from co-authors or collaborators to post my work on publicly accessible web sites. (item3) • I cannot publish my work if I post it on publicly accessible web sites before publication. (item4)
	Additional time and effort	• Posting my materials on publicly accessible web sites takes time away from my research and writing. (item5) • Additional time and effort is required to make my materials publicly accessible on the Internet. (item6) • It is difficult to learn how to enter the required data (e.g., title, author, date etc.) with my deposit. (item7)
Extrinsic benefit factors	Accessibility	• Posting my research work on publicly accessible web sites will increase the chance to communicate my research findings to peers. (item8) • Materials on publicly accessible web sites are more easily accessible through Internet search engines. (item9) • Materials on publicly accessible web sites are not preserved in perpetuity. (item10)
	Publicity	• Posting my materials on publicly accessible web sites will enlarge the readership of the materials. (item11) • Posting my research work on publicly accessible web sites will increase the potential impact of my work. (item12) • Posting my research work on publicly accessible web sites allows for earlier dissemination of my research findings. (item13)
	Trustworthiness	• I trust the quality of materials on publicly accessible web sites from authors employed by prestigious institutions. (item14) • I trust the quality of materials on publicly accessible web sites from well-known researchers in my field. (item15) • I trust the quality of peer-reviewed articles on publicly accessible web sites. (item16)
	Professional recognition	• Posting my research work on publicly accessible web sites will increase my visibility within the discipline(s) to which I belong. (item17) • Materials on publicly accessible web sites will be cited more frequently. (item18)

	Academic reward	• My university will accept research work on publicly accessible web sites as an alternative to publication for tenure and promotion. (item19) • Posting my work on publicly accessible web sites will adversely affect my chances of tenure/promotion. (item20) • Posting my research work on publicly accessible web sites will adversely affect my chances of attaining research grants. (item21)
Intrinsic benefit factor	Altruism	• I will continue posting my work on publicly accessible web sites even if others in my field do not. (item22) • I support the principle of open access (free and unrestricted access to research materials) for all users. (item23) • Posting my materials on publicly accessible web sites will help other researchers build on my research findings. (item24) • Posting my materials on publicly accessible web sites allows other scholars to access those that they could not otherwise use. (item25)
Contextual factors	Trust	• If I post my materials on publicly accessible web sites, readers may plagiarize or fail to cite my work. (item26) • If I post my materials on publicly accessible web sites, the integrity of my work will be compromised. (item27) • Materials on publicly accessible web sites are not maintained securely. (item28)
	Self-archiving culture	• In my field, it is common for researchers to post their pre- or post-refereed articles on publicly accessible web sites. (item29) • In my department, it is common for faculty and students to create and share working papers or technical reports. (item30)
	Influence of external actors	• My decision to make, (or not to make) my materials publicly accessible on the Internet was influenced by my co-authors or collaborators. (item31) • My decision to make, (or not to make) my materials publicly accessible on the Internet was influenced by my grant-awarding body. (item32) • My decision to make (or not to make) my materials publicly accessible on the Internet was influenced by my university or department. (item33)

Some of the items were adapted from those used in studies based on survey methodology (Gadd et al., 2003a; Kankanhalli et al., 2005; Swan and Brown, 2005), whereas others I have newly developed based on the literature reviewed. For example, items regarding the 'influence of external actors' were modified from those employed in the study of Swan and Brown (2005). Although they did not categorize those items as the influence of other actors, those statements reflect the concept, which Kling et al. (2003) suggested in their socio-technical network model. In addition, items concerning trustworthiness were created based on the concept proposed by Kling and McKim (1999) and Cronin (2005). These studies mentioned that readers normally assessed the quality of research materials based on (1) whether or not those were peer-reviewed; (2) reputations of authors; and (3) the prestige of institutions by which authors employed. Using these dimensions, I developed three items measuring the degree to which respondents would trust the quality of research work.

Pilot Survey

A small-scale pilot survey was conducted in July-August 2006, in order to determine the reliability of scale questions, and to check whether the administrative procedure of survey distribution worked smoothly. I made a random sample of 40 professors from 12 prototypical disciplines in a Carnegie Doctoral-granting University. The disciplines included (1) Science (Physics, Mathematics, and Molecular Biology); (2) Engineering (Computer Science, Electrical and Computer Engineering, and Mechanical Engineering); (3) Social Science (Economics, Sociology, and Psychology); and (4) Humanities (Art, English, and History). I randomly selected 3 or 4 professors in each discipline.

I distributed paper invitation letters, each of which enclosed $5 gift card. An individual letter included a link to the online survey using SurveyMonkey, as well as an

ID assigned to each professor to identify who responded the survey. Once the letters were delivered to the sample, I sent initial follow-up e-mails including the link to the online survey. After two weeks, second follow-up e-mails were distributed.

Out of 40 professors, 20 (50.0%) completed the survey. Twelve (60.0%) professors have self-archived their research materials. Posting refereed, published articles on personal web sites was typical self-archiving behavior. Yet, no respondents contributed to their university IR. The collected data of scale questions were used to calculate Cronbach's alpha to determine reliability as described in the following section.

Validation

Reliability

According to DeVellis (2003), scale reliability is defined as "the proportion of variance attributable to the true score of the latent variable" (p.27). DeVellis suggests that if items of a scale have a strong association to their latent variable, the items will have a strong relationship to each another. Although it is not feasible to directly observe the relationship between items and the latent variable, it is certainly possible to see whether items are inter-correlated. The internal consistency reliability, therefore, is represented as the extent to which the items are highly correlated to one another. While there are various ways of computing reliability, the most common method for determining internal consistency of a scale is to compute Cronbach's coefficient alpha. Alpha indicates the proportion of the total variance of a scale ascribed by the true score of a latent variable. Therefore, alpha divides the total variance into actual variation due to the latent variable, and error.

A coefficient alpha of **.**70 or higher is considered to be the acceptable level of internal consistency among items in most social science research situations, although the minimum threshold of alpha is .50. Based on a pilot survey data, 6 factors (publicity,

trustworthiness, professional recognition, altruism, trust, and self-archiving culture) showed alpha values higher than .70, and the remaining 5 factors (copyright concerns, additional time and effort, accessibility, academic reward, and influence of other actors) satisfied the minimum threshold.

Validity

An issue of validity is to determine "the adequacy of a scale as a measure of a specific variable" (DeVellis, 2003, p.49). DeVellis suggests three types of validity: (1) content validity; (2) criterion-related validity; and (3) construct validity.

Content validity is defined as the degree to which a set of items manifests a content domain. In a theoretical sense, content validity of a scale can be accomplished when its items are randomly selected from the universe of appropriate items. However, if the scale measures beliefs or opinions, such a comprehensive pool of items cannot exist. In this case, a common method is to have colleagues, familiar with the context of research, review an initial list of items and ask them to suggest which items should be excluded or added. The survey instrument used in the present study has already been reviewed by a selection of professors with expertise in user behavior or digital preservation. Based on their suggestions, some Likert-scale questions were omitted, and others were revised or newly added.

Criterion-related validity is determined by empirical association between a scale of items, and some criterion in a real world. DeVellis mentions that criterion-related validity is not related to theoretical issues, but to practical concerns with predicting a certain phenomenon. In this sense, criterion-related validity often refers to predictive validity. In addition, other types of criterion-related validity exist, depending on whether the criterion follows, precedes or co-occurs with the measurement pertaining to the questions. Since some of the scale questions in the present study are based on empirical research on self-archiving practices, the collected data showed similar results to those of

the empirical literature. The pilot survey, however, collected only 20 responses and with this small set of survey data, it was difficult to determine criterion-related validity.

Construct validity is closely related to theoretical discourse, regarding the relationship between variables. In theory, if construct A has a relationship to other constructs, a scale designed to measure construct A should reflect a similar association to measures of other constructs. This study employs socio-technical network models and social exchange theory to identify conceptual relationships between self-archiving behavior and various factors displayed in Table 3.1. Construct validity represents the degree to which the theoretical relationship is translated into actual associations between measurements of dependent variables and those of independent variables. Confirmatory Factor Analysis (CFA) is a method to determine construct validity. However, since the pilot survey data was too small (n=20), the result of CFA was not meaningful.

The survey instrument, therefore, was reviewed by experts to establish content validity prior to survey data gathering, although criterion-related validity and construct validity were not determined. Yet, construct validity was established based on CFA of actual survey data (see Appendix D).

Survey Data Collection

As described in Pilot Survey section, the initial contact to the 1,500 surveyed sample was done by sending paper invitation letters via postal mail (Appendix B). Each invitation letter provided a link to the web survey and an ID number for each sample. In addition, $2 bill was enclosed in each letter as an incentive. Birnholtz et al. (2004) found that online surveys preceded by paper invitations with a $5 cash incentive, generated a significantly higher response rate (57%) than those online surveys in two other conditions: (1) by sending a paper invitation with $5 online gift certificate code; and (2) an e-mail invitation providing a $5 online gift certificate code. Response rates under

those conditions were found to be the same (36%). This study indicates that cash incentives have a significant impact on response rates compared to the same value of online gift certificates.

The survey invitation letters with $2 incentive cash were distributed to 1500 professors at 17 universities from October 13 to 27 in 2006. The sample was divided into three groups to facilitate mailing procedure in a convenience manner: (1) 1st group including 254 professors from 2 universities; (2) 2nd group including 445 professors from 5 universities; (3) 3rd group including 801 professors from 10 universities. The letters were mailed to the 1st group in October 13, to the 2nd group in October 20, and to the last group in October 27. After one week from each of the mailing dates, I sent initial follow-up e-mails, each of which enclosed an online survey link. After that, those follow-up e-mails were delivered twice in two-week interval. The survey was closed ten days after sending the last follow-up e-mails. Table 3.4 presents dates of sending invitation letters and follow-up e-mails, and closing the survey.

Table 3.4. Dates of Distributing Survey, Follow-up E-mails, and Closing the Survey

	Mailing invitation letters	**1st follow-up e-mail**	**2nd follow-up e-mail**	**3rd follow-up e-mail**	**Closing the survey**
1st group	Oct. 13	Oct. 20	Nov. 3	Nov. 17	Nov. 27
2nd group	Oct. 20	Oct. 27	Nov.10	Nov. 24	Dec. 4
3rd group	Oct. 27	Nov. 3	Nov.17	Dec.1	Dec. 11

The number of survey responses originally submitted was 736, while it included several totally empty responses. I excluded all the substantially incomplete responses and the resulting number of survey responses used for the analysis was 684 (45.6%).

Survey Sample vs. Respondents

The survey sample consisted of two groups: (1) IR contributors defined as professors whose materials are deposited in their university IRs; (2) IR non-contributors, the remaining number of professors in the 17 universities. IR contributors were located

by identifying author lists in the sample universities' IRs and then searching assistant, associate, and full professors via university online directories. As a result of the search process, I found 621 IR contributors. Almost 90% of these individuals were tenured professors; 47% were from engineering fields. In order to balance the distribution of the entire sample by rank and disciplines, I oversampled assistant professors among IR non-contributors and attempted to include more professors from science, social science, and humanities.

In fact, IR non-contributors constituted most faculty members in the 17 universities in various disciplines. Since identifying all the faculty members would take a vast amount of time and effort, I focused instead on three prototypical disciplines in each of the four areas: Science, Engineering, Social science, and Humanities (See Table 3.5).

Table 3.5. Twelve Disciplines of IR non-contributor Sample

Area	Disciplines
Science	Mathematics, Molecular Biology, and Physics
Engineering	Computer Science, Electrical and Computer Engineering, and Mechanical Engineering
Social Science	Economics, Psychology, and Sociology
Humanities	Art, English, and History

In those 12 disciplines at the 17 universities, I randomly selected professors by looking at the faculty lists of department web sites and sampling them based on random numbers generated by MS-Excel. Consequently, 879 professors were sampled as IR non-contributors. Therefore, the entire sample included 1,500 professors (621 IR contributors and 879 non-contributors).

Rank

Out of 1,500 sampled professors, 459 (30.6%) were assistant professors, whereas the rest were tenured professors including associate, full, and distinguished professors (See Table 3.6). Rank was identified through descriptions of the professors on their departmental web pages.

Table 3.6. Sample by Rank

	IR contributors		IR non-contributors		Total	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Assistant Professor	63	10.1	396	45.1	459	**30.6**
Associate Professor	173	27.9	170	19.3	343	**22.9**
Full Professor	385	62.0	296	33.7	681	**45.4**
Distinguished Professor	0	0.0	17	1.9	17	**1.1**
Total	621	100.0	879	100.0	1500	100.0

Out of 1,500 sampled professors, 684 (45.6%) either entirely or substantially completed the questionnaire. Of the 621 IR contributors, 269 responded, whereas 415 answered the survey among 879 IR non-contributors. Table 3.7 shows the distribution of respondents by rank.

Table 3.7. Respondents by Rank

	IR contributors		IR non-contributors		Total	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Assistant Professor	33	12.3	186	44.8	219	**32.0**
Associate Professor	80	29.7	85	20.5	165	**24.1**
Full Professor	130	48.3	122	29.4	252	**36.8**
Distinguished Professor	26	9.7	22	5.3	48	**7.0**
Total	269	100.0	415	100.0	684	100.0

More respondents identified themselves as distinguished professors and the respondents' ranks in Table 3.7 are based on their survey answers, while ranks associated with the initial sample (Table 3.6) are based on departmental web pages. Those web pages did not always provide the most updated titles of professors once they were promoted. I occasionally found that some assistant professors in the sample identified themselves as associate professors in survey answers. Overall, the distribution of respondents by rank was almost proportionate to the sample distribution.

Disciplines

Professors sampled as IR contributors were from a variety of disciplines. Those selected as IR non-contributors belonged to twelve disciplines in four areas (See Table 3.5). As seen in Table 3.8, I selected more professors in Science, Social Science, and the

Humanities than in Engineering to form the sample of IR non-contributors, since the majority of IR contributors were engineering professors.

Table 3.8. Sample by Disciplines

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Science	77	12.4	242	27.5	319	**21.3**
Engineering	294	47.3	171	19.5	465	**31.0**
Social Science	211	34.0	225	25.6	436	**29.1**
Humanities	39	6.3	241	27.4	280	**18.7**
Total	621	100.0	879	100.0	1500	100.0

The distribution of respondents by disciplines is presented in Table 3.9 below. It is almost proportionate to the sample distribution by disciplines.

Table 3.9. Respondents by Disciplines

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Science	34	12.6	127	30.6	161	**23.5**
Engineering	114	42.4	78	18.8	192	**28.1**
Social Science	96	35.7	110	26.5	206	**30.1**
Humanities	25	9.3	100	24.1	125	**18.3**
Total	269	100.0	415	100.0	684	100.0

As mentioned earlier, respondents from the IR contributor group belonged to diverse disciplines (see Appendix E). Among respondents from scientific disciplines, the most were from Mathematics. In Mathematics 63 professors were sampled (17 IR contributors and 46 non-contributors), followed Physics professors with 48 professors (4 IR contributors and 44 non-contributors), and 41 Biology professors (4 IR contributors and 37 non-contributors). The specializations of the Engineering respondents varied. The majority was from Electrical and Computer Engineering (32 IR contributors and 34 non-contributors), followed by 45 Computer Science professors (25 IR contributors and 20 non-contributors), and 43 Mechanical Engineering professors (19 IR contributors and 24 non-contributors). The Engineering respondents who were IR contributors tended to be involved in several other disciplines, such as horticultural science, chemical engineering, or civil and environmental engineering.

Most of the Social Science professors (54) were from Economics (27 IR contributors and 27 non-contributors), followed by 50 Psychology (9 IR contributors and 41 non-contributors), and 48 Sociology (6 IR contributors and 42 non-contributors). Twenty-five were from business and management and the remainder were scattered among a variety of areas such as Education, Public Policy, and Political Science.

The least number of respondents came from the Humanities. Among IR contributors in the Humanities, Linguistics professors led the way (n=8). However, when considering both IR contributors and non-contributors, the majority was from English (41 professors; 1 IR contributor and 40 non-contributors), followed by 40 History (3 IR contributors and 37 non-contributors), and 25 Art (1 IR contributor and 24 IR non-contributors).

Universities

The sample included current professors at 17 Carnegie Doctorate-granting Universities. All these universities implemented DSpace Institutional Repositories (IRs), although the maturity of the IR implementation differs in each university.

Table 3.10 shows the distribution of sampled professors by disciplines. The names of universities were presented as letter **A** to **Q** based on the total number of titles at each university's IR in February 2006, the date on which I finished identifying faculty IR contributors via online directories of each university (See Figure 3.1).

Out of the IR contributor sample, 200 (32.2%) were from **A** University, which contained the largest number of titles in its IR (n=19,046). It should be noted that the number of faculty members contributing to the IR was not always consistent with the size of IR collection. For example, **C** University had the third largest IR collection; however, two faculty IR contributors were identified.

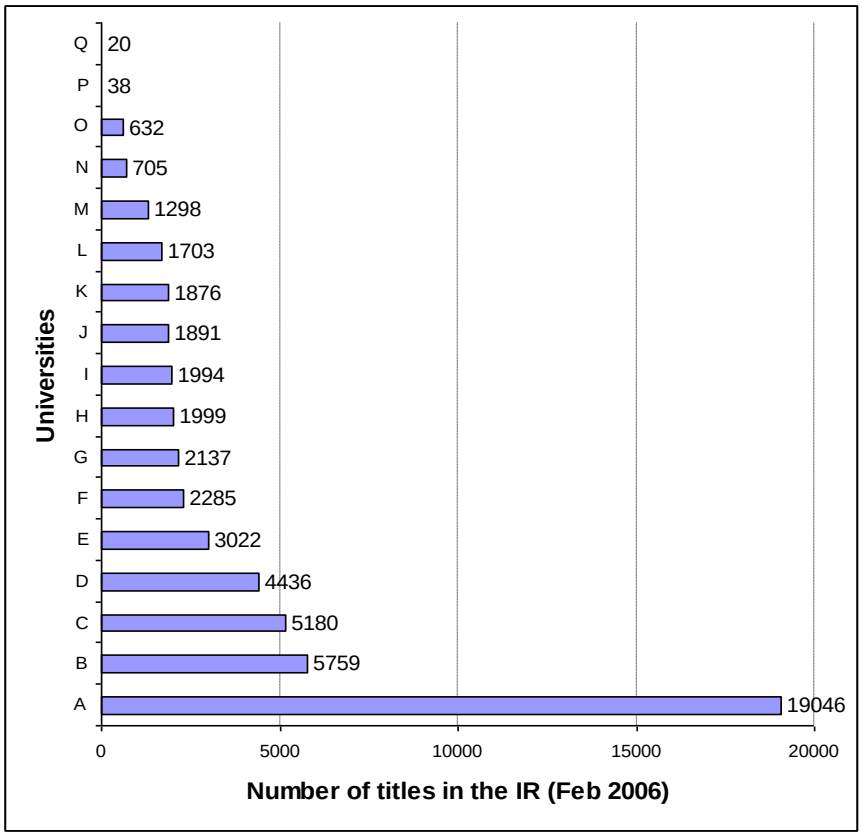

Figure 3.1. Number of titles in each university's IR (in February 2006)

Table 3.10. Sample by University

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
A	200	32.2	42	4.8	242	16.1
B	90	14.5	40	4.6	130	8.7
C	2	0.3	57	6.5	59	3.9
D	12	1.9	57	6.5	69	4.6
E	3	0.5	53	6.0	56	3.7
F	25	4.0	56	6.4	81	5.4
G	8	1.3	51	5.8	59	3.9
H	47	7.6	54	6.1	101	6.7
I	18	2.9	57	6.5	75	5.0
J	35	5.6	42	4.8	77	5.1
K	21	3.4	50	5.7	71	4.7
L	9	1.4	52	5.9	61	4.1
M	63	10.1	52	5.9	115	7.7
N	3	0.5	51	5.8	54	3.6
O	68	11.0	56	6.4	124	8.3
P	10	1.6	55	6.3	65	4.3
Q	7	1.1	54	6.1	61	4.1
Total	621	100.0	879	100.0	1500	100.0

Table 3.11 presents the distribution of respondents by universities. The number of respondents from **A** University was 69 (10.1%), which was much smaller than the proportion of the sample from the university (16.1%). Also, the proportions of respondents in five universities (**B**, **C**, **E**, **L**, and **M**) were slightly smaller than those of the sample. However, the proportions of respondents in ten universities (**D**, **F**, **G**, **H**, **J**, **K**, **N**, **O**, **P**, and **Q**) were slightly greater than those of the sample. The percentage of respondents in **I** University was the same as that of the sample. Despite of the discrepancies, the distributions of the sample and respondents by university were proportionate overall.

Table 3.11. Respondents by University

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
A	53	19.7	16	3.9	69	10.1
B	34	12.6	15	3.6	49	7.2
C	0	0.0	25	6.0	25	3.7
D	7	2.6	30	7.2	37	5.4
E	1	0.4	19	4.6	20	2.9
F	20	7.4	18	4.3	38	5.6
G	5	1.9	24	5.8	29	4.2
H	29	10.8	19	4.6	48	7.0
I	11	4.1	23	5.5	34	5.0
J	19	7.1	26	6.3	45	6.6
K	7	2.6	28	6.7	35	5.1
L	3	1.1	24	5.8	27	3.9
M	25	9.3	22	5.3	47	6.9
N	2	0.7	32	7.7	34	5.0
O	40	14.9	33	8.0	73	10.7
P	8	3.0	31	7.5	39	5.7
Q	5	1.9	30	7.2	35	5.1
Total	269	100.0	415	100.0	684	100.0

Interview Data Collection

The survey provided a question asking whether respondents would not be willing to participate in 30-minute follow-up phone interviews. Those who did not answer the

question were considered potential interviewees whom I could contact, and 151 respondents were found not to check the question. Along with them, I created a list of 64 respondents including 53 self-archivers and 11 non-self-archivers. Among the 53 self-archivers, 26 identified themselves to be IR contributors, whom I assumed would be more interested in self-archiving than other respondents.

For recruiting interviewees, the 64 respondents were contacted by e-mail and 26 agreed to participate in the phone interviews. Out of the 26 interviewees, 18 were self-archivers and 9 were non-self-archivers. After that, I contacted the remaining 87 respondents, consisting of 61 self-archivers and 26 non-self-archivers. As a result, fourteen respondents, all who have self-archived were interviewed. In sum, I performed forty-one phone interviews in March-May 2007.

It took 20 minutes on average to conduct phone interviews. Interview protocols were employed for all 41 interviews. The interview protocol for self-archivers was different from that for non-self-archivers (see Appendix C). All interviews were digitally recorded and transcribed by a professional transcriber.

The collected survey data was analyzed in using Stata 8.2 and the interview data was coded and analyzed by employing QSR Nvivo7. The following chapter describes findings from the analyses of the qualitative and quantitative data.

CHAPTER 4

RESULTS

The present study investigates a wide range of self-archiving behaviors by university faculty members. Two types of data were collected: survey and interview. The survey data captured various ways of making several types of research work publicly accessible on the web and participants motivations for these activities. The interviews probe more deeply into subjects' motivation and reasoning behind self-archiver behaviors. This chapter initially discusses survey findings then discusses a series of interviews that build on these results. In discussing the survey, I will first illustrate several aspects of self-archiving practices including self-archived content types, self-archiving venues, frequency of self-archiving, IR contribution, and respondents' intention of self-archiving in the future. After that, results of both Ordinary Least Squares (OLS) and logistic regression analysis are reported to explore significant factors affecting percentages of self-archived research work, and whether or not respondents had contributed to their university's IR.

After reporting these survey data, I will discuss the interview data analyses, which provide diverse perspectives of self-archiving practices and interviewees' rationale for participating or not participating in such practices. I will also triangulate results of survey and interview data, determining consistencies or differences between the two types of data.

Survey Data Analysis

Self-archiving Practices

The present study examined to what extent the 684 faculty respondents have been involved in self-archiving. Self-archiving is defined as making research publicly available on the Web. Research includes various versions of research papers, such as pre-refereed articles, publishers' PDF versions of refereed articles, authors' final versions of refereed articles, and unrefereed articles (i.e., technical reports or working papers), book chapters, as well as research data sets. These types of research materials are common across multiple disciplines. In addition, the survey respondents were professors at large research universities, and therefore, it was assumed that most respondents had produced these types of output through their research process.

Self-archivers

Out of the 684 faculty respondents, 480 (70.2%) had made research materials publicly available on the Internet. Tables 4.1 and 4.2 present ranks and disciplines of these self-archivers, respectively.

Table 4.1. Self-archivers by Rank

	Freq.	*Percent*
Assistant Professor	145	30.2
Associate Professor	121	25.2
Full Professor	176	36.7
Distinguished Professor	38	7.9
Total	480	100.0

Table 4.2. Self-archivers by Disciplines

	Freq.	*Percent*
Science	123	25.6
Engineering	160	33.3
Social Science	146	30.4
Humanities	51	10.6
Total	480	100.0

Within the 684 respondents, the distribution of self-archivers by rank and disciplines was also examined. Table 4.3 indicates the academic rank of the self-archivers in terms of overall number and percentages. The percentage of assistant professors is 66.2%, slightly smaller than the total proportion of self-archivers (70.2%). Overall, however the percentage of self-archivers by rank is close to the total percentage.

Table 4.3. Academic Rank of Respondents

Academic Rank	*No. of self-archivers*	*No. of respondents*	*Percent of self-archivers*
Assistant Professor	145	219	**66.2**
Associate Professor	121	165	**73.3**
Full Professor	176	252	**69.8**
Distinguished Professor	38	48	**79.2**
Total	480	684	**70.2**

Table 4.4 presents percentages of self-archivers by disciplinary area. The proportion of engineering professors is 83.3%, which is greater than the total percentage of self-archivers. The proportion of humanities professors, however, is 40.8%, far lower than the total percentage. This indicates that humanities professors are less likely to self-archive their research materials than professors were in other disciplines.

Table 4.4. Proportion of Self-archivers to Total Respondents by Disciplines

	No. of self-archivers	*No. of respondents*	*Percent of Self-archivers*
Science	123	161	**76.4**
Engineering	160	192	**83.3**
Social Science	146	206	**70.9**
Humanities	51	125	**40.8**
Total	480	684	**70.2**

More than half (51.9%) of the faculty self-archivers have made their research papers or books publicly available on the web for over five years; 23.5% have done self-archiving for 1-3 years. Twelve respondents did not answer the question (See Table 4.5).

Table 4.5. How Long Respondents Had Self-archived

	Freq.	*Percent*
1 year or less	17	3.5
1-3 years	113	23.5
4-5 years	89	18.5
More than 5 years	249	51.9
N/A	12	2.5
Total	480	100.0

Table 4.6 also indicates faculty who actually post research papers or books on publicly accessible web sites. Out of the 480 self-archivers, 399 (83.1%) indicated they actually self-archived themselves. Student assistants, department staff, and collaborators self-archived for them less frequently. Faculty also identified others who posted their content online including publishers, web designers that respondents hired, conference organizers and research center staff.

Table 4.6. People who are Actually Involved in Self-archiving

	Freq.	*Percent*
Myself	399	83.1
Students or assistants	143	29.8
Department staff	117	24.4
Collaborators	100	20.8
Library staff	38	7.9
Other	21	4.4

Content Types Self-archived

The 480 faculty self-archivers made various versions of research papers publicly accessible on the Internet. The types of papers included (1) pre-refereed drafts; (2) publishers' PDFs of refereed articles; (3) authors' final versions of refereed articles; (4) unrefereed articles, such as technical reports or working papers; (5) book chapters. Respondents could indicate multiple types when answering this question. Table 4.7 shows how many self-archivers have made each type of research papers and book chapters publicly accessible on the Web.

Table 4.7. Research Papers and Book Chapters Posted by Self-archivers

	Freq.	*Percent*
Final versions of refereed articles	331	69.0
Pre-refereed articles	318	66.3
Unrefereed articles	310	64.6
Publishers' PDFs	273	56.9
Book chapters	153	31.9

Overall, book chapters were much less frequently self-archived. One self-archiver (ID 1327) made a comment that he was more interested in posting his articles than books because he "still prefers that people buy the books!" Among the other types, publishers' PDFs of refereed articles were slightly less often self-archived than pre-refereed, unrefereed, or final versions of refereed articles.

Considering the fact that many publishers do not allow authors to post publishers' PDF versions on the web, it was interesting to learn that 273 (56.9%) self-archivers had made the PDF versions publicly available on the Internet. One of the self-archivers (ID 1537) clearly recognized the copyright issue and mentioned in the survey, "I respect copyright owners and will not put up PDFs of published papers." Another self-archiver (ID 2363) even commented, "I suspect that most faculty who post articles on their public web sites do not know that they are violating copyright laws." Although this comment was not always true, depending on publishers in different fields, several self-archivers stated that they were not clear how to deal with copyright when posting their articles. Yet, one respondent (ID 1247) even mentioned that he ignored publishers' policies that did not allow for self-archiving because those were "silly".

Interestingly, 19 (4.0%) out of the 480 self-archivers had not self-archived any of their research papers or books. Eight were professors of Art and some of whom commented that papers or books did not take into account creative activity such as dance, music, visual arts, or theater, even though they posted all their artwork on the Internet. One respondent stated that "in my field of theatrical scenic design, I have not utilized this practice and am not aware of its general use. The one exception would be the similar

practice of posting digital design portfolios online" (ID 2149). What they considered research work was very different from those in other disciplines and they did not see IRs as accommodating these genres. The rest of the professors mentioned that they posted research materials separate from papers or books. For example, one mechanical engineering professor had posted movie files so that reviewers of proposals or manuscripts might have access to those files. Moreover, two biology professors had self-archived software and they considered it research work.

In addition, Figure 4.1 presents percentages of research papers and data sets that the 480 self-archivers made publicly available on the Web in the last five years.

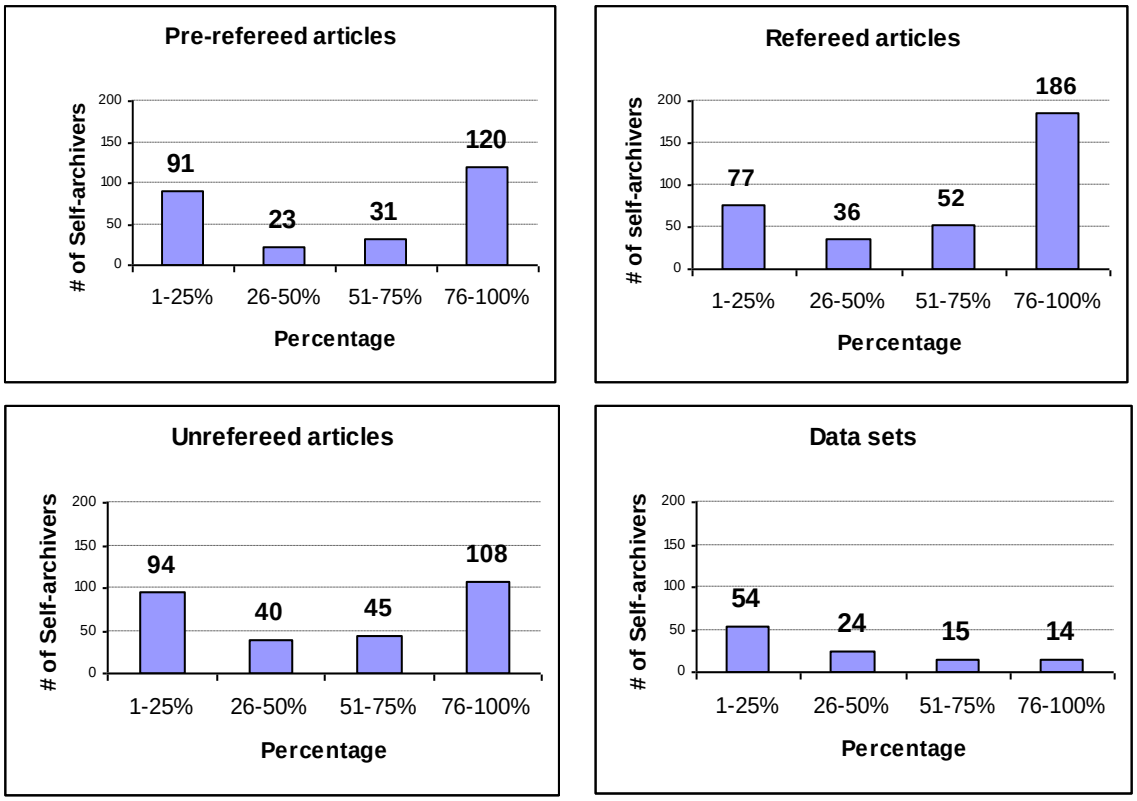

Figure 4.1. Percentages of Research Papers and Data sets Self-archived

Self-archivers usually make a greater proportion of refereed articles publicly accessible on the Web than any other type of research work. Fifty-two (10.8%) self-archivers post 51-75% of their refereed articles, and 186 (38.8%) post 76-100% of refereed articles on publicly available web sites. The fewest number of self-archivers post

data sets on publicly available web sites. The majority of those who self-archive data sets have made 1-25% of their research data publicly accessible on the Internet.

Other than research papers, books, or data sets, self-archivers have made diverse kinds of research or teaching materials publicly accessible on the Web. Figure 4.2 shows the distribution of other content types self-archived.

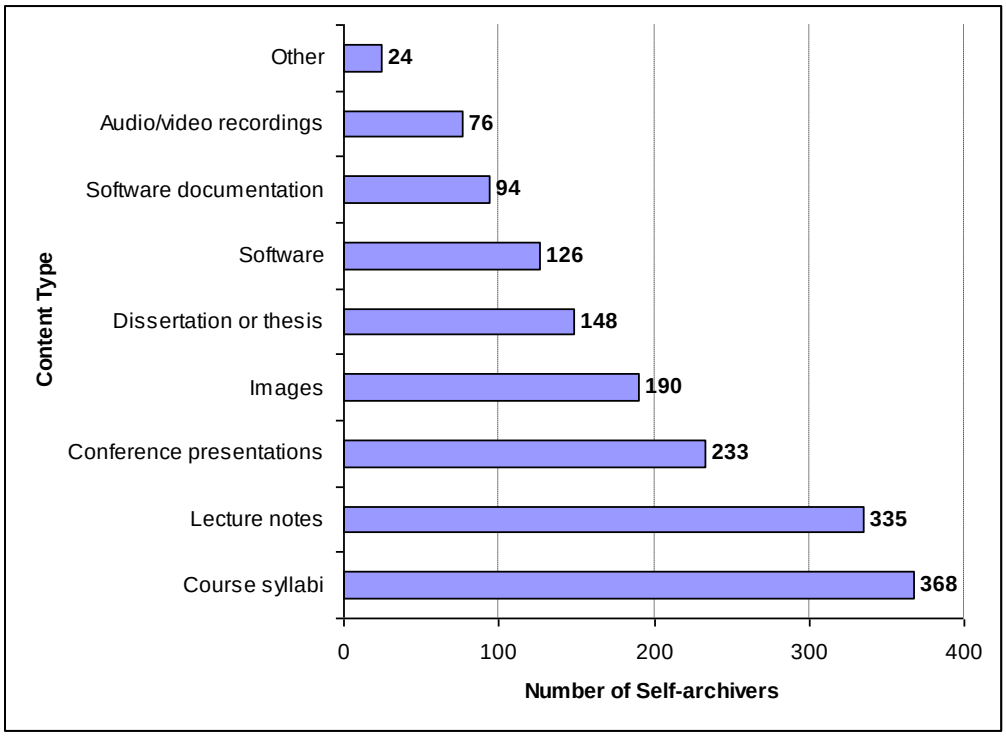

Figure 4.2. Other Content Types Self-archived

Three-hundred sixty eight (76.7%) self-archivers have posted course syllabi on publicly accessible web sites, and 335 (69.8%) have made lecture notes, handouts, and assignments publicly available on the Internet. Conference presentations are another frequently self-archived content type. Almost half of the self-archivers (48.5%) made these publicly accessible on the Web. Software is one of the non-traditional types of scholarly contents and 126 (26.3%) have posted it on publicly accessible web sites. Additionally, 190 (39.5%) and 76 (15.8%) have made images and audio/video recordings publicly available on the Internet, respectively. Twenty-seven (5.6%) self-archivers have made none of these content types publicly accessible on the Web.

In sum, it was refereed articles that were the most frequently self-archived content type. The largest number of survey respondents self-archived their own versions of refereed articles. However, almost 57% of survey respondents posted publishers' PDFs on publicly accessible web sites, although it is typically not allowed by publishers. Survey respondents have also self-archived greater percentage of refereed articles than those of other research materials. Yet, faculty members made a variety of content types other than research papers publicly accessible on the Web, such as course materials and conference presentations.

Venues for Self-archiving

The 480 faculty self-archivers primarily employed the following five open access venues for self-archiving: (1) personal web pages; (2) research group/center/lab web sites; (3) departmental web sites; (4) disciplinary repositories; and (5) institutional repositories (IRs). Table 4.8 presents how many self-archivers have used such venues for self-archiving. Respondents were allowed to choose multiple venues.

Table 4.8. Venues used by Self-archivers

	Freq.	*Percent*
Personal web pages	320	66.7
Research group web sites	247	51.5
Departmental web sites	200	41.7
Disciplinary repositories	136	28.3
IRs	109	22.7

As seen in the table above, personal web pages are the most popular venue for self-archiving. Research group/center/lab web sites or departmental web sites were not as frequently used as personal web pages, although many more self-archivers used such web sites than either disciplinary or institutional repositories. Disciplinary repositories only exist in certain disciplines, for example, arXiv.org in Physics. Therefore, disciplinary repositories were not used widely across the disciplines. Although IRs were implemented in all the universities, by which the self-archivers were employed as

professors, the fewest number of the self-archivers contributed their research papers or books to IRs in their universities.

Frequency of Self-archiving

Faculty self-archivers tended to post refereed articles on publicly accessible web sites, particularly their own personal web pages. Figure 4.3 shows the frequency of self-archiving research papers or books on personal web pages in the past 3 years. At most, 224 (46.6%) of the self-archivers have posted one of the five most common types of research materials – pre-refereed articles, final versions of refereed articles, publishers' PDFs, unrefereed articles, and book chapters - on personal web pages. When asked how many times survey respondents posted the content types in the last 3 years, survey respondents chose one of the four categories of frequencies: (1) once; (2) 2-3 times; (3) 4-5 times; (4) over 6 times. I categorized over 6 times as very frequently, 4-5 times as moderate, 2-3 time as low, and once as sporadic.

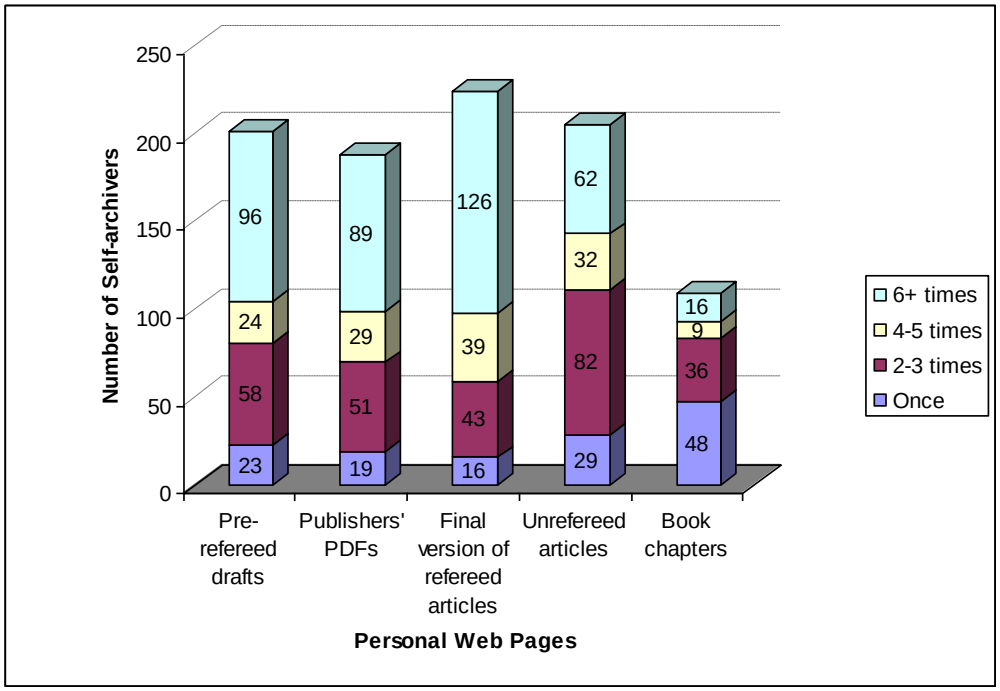

Figure 4.3. Frequency of Self-archiving on Personal Web Pages

As seen in Figure 4.3, final versions of refereed articles are self-archived most frequently. 126 (56.2%) out of 224 self-archiving respondents who have self-archived final versions of refereed articles, reported posting them on personal web pages over 6 times. However, the remaining 98 respondents (43.8%) self-archived final versions of refereed articles on personal web pages 5 times or less in the past 3 years, which indicated moderate or low frequency of self-archiving. This finding suggested that self-archiving was not a regular and frequent activity that faculty members conducted. Similarly, this pattern of self-archiving frequency was found in self-archiving other research content in different web venues.

In addition, around 30% of self-archivers have posted research papers or books on research group/center/lab web sites in the past 3 years. Similar to self-archiving on personal web pages, final versions of refereed articles were posted on research group web sites. As seen in Figure 4.4, 74 (48.7%) out of 152 respondents who posted final versions of refereed papers on research group web sites, self-archived them over 6 times. However, the rest of the respondents (n=78, 51.3%) self-archived the refereed articles 5 times or less.

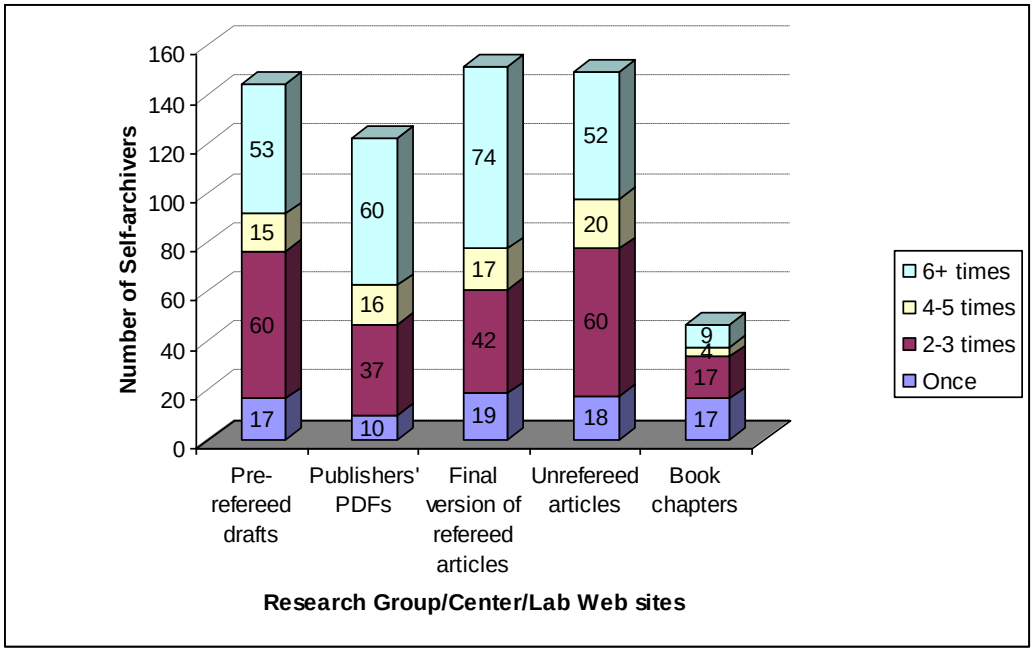

Figure 4.4. Frequency of Self-archiving on Research Group/Center/Lab Web Sites

Figure 4.5 presents that more self-archivers have posted pre-refereed drafts than other types of research papers on departmental web sites. Forty-two (34.4%) out of 122 respondents who posted pre-refereed drafts on departmental web sites, reported self-archiving them over 6 times. Yet, 80 (65.6%) respondents self-archived pre-refereed drafts 5 times or less. The percentage of respondents who self-archived infrequently (5 times or less) on departmental web sites was much greater than that of respondents who self-archived more frequently (over 6 times).

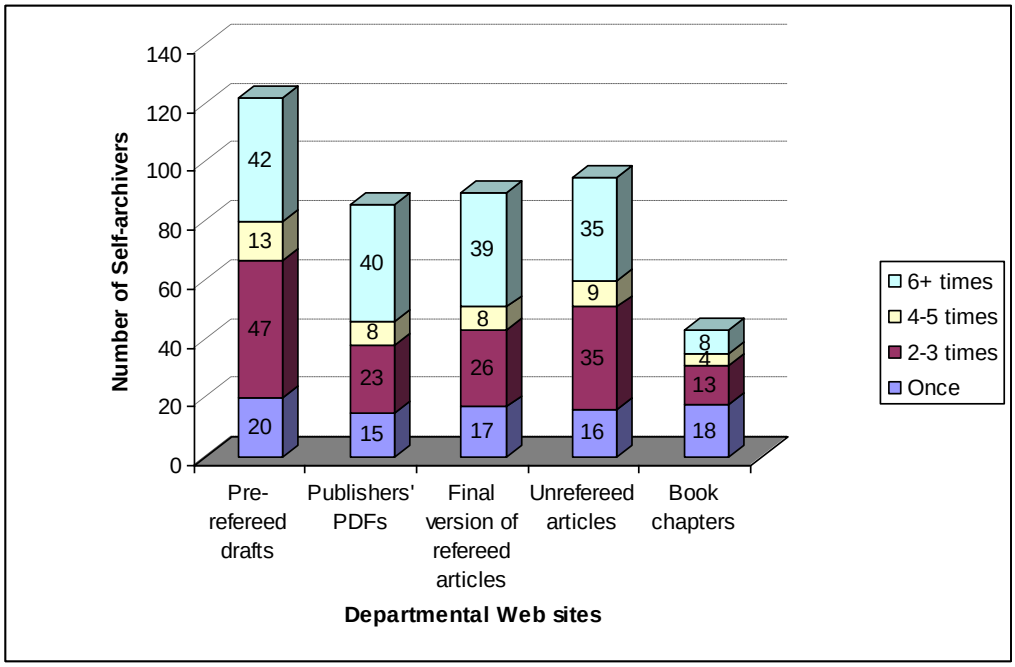

Figure 4.5. Frequency of Self-archiving on Departmental Web Sites

Less than 25% of self-archivers posted pre-refereed drafts on disciplinary repositories (see Figure 4.6). It is also interesting to note that 31 (22.8%) out of the 136 self-archivers in disciplinary repositories were from the Social Sciences. Of the 31 social scientists, 16 were in Economics and 9 in Business Administration. Especially in Economics, the research paper series in the Economics Research Network (ERN) was commonly used by researchers in the field (David and Connolly, 2007). Therefore, the majority of professors in social science who deposited pre-refereed drafts in disciplinary repositories were from Economics and related fields.

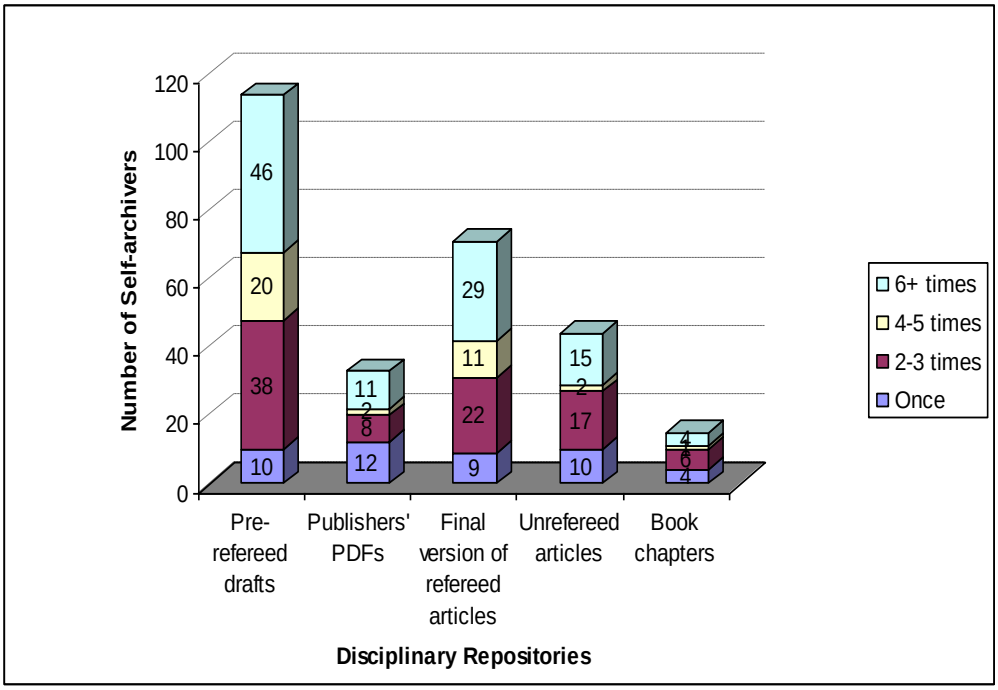

Figure 4.6. Frequency of Self-archiving on Disciplinary Repositories

Approximately 15% of self-archivers have contributed research papers or books to institutional repositories (IRs).

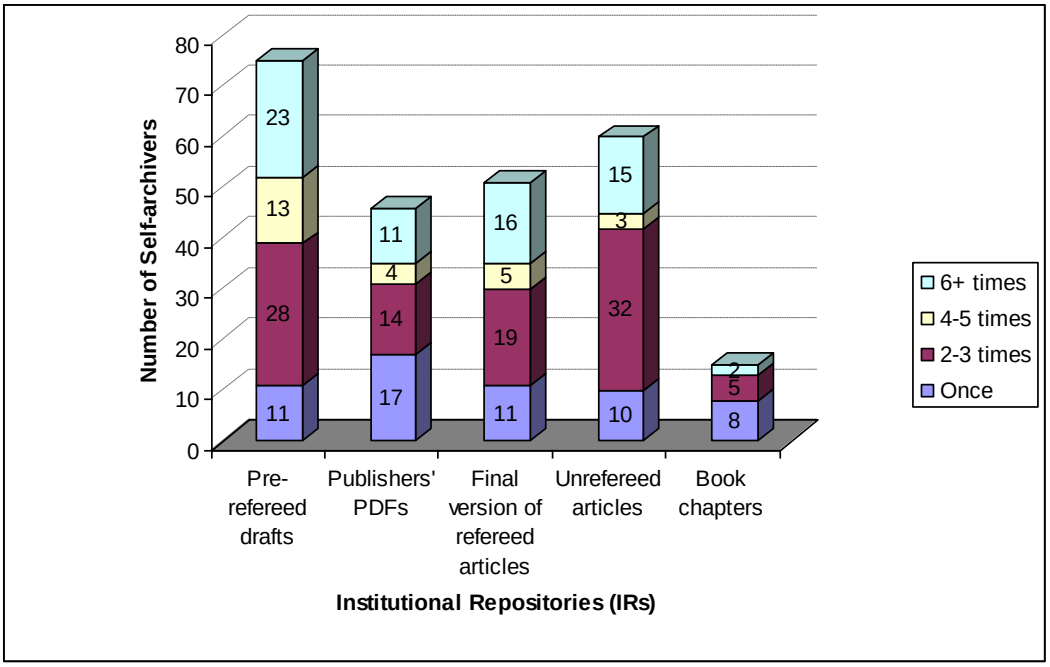

Figure 4.7. Frequency of Self-archiving on Institutional Repositories (IRs)

Among the various self-archiving venues, IRs were marginally used by faculty self-archivers, despite the fact that all the universities in this study provided IR services. Figure 4.7 shows the frequency of contributing research papers or books to IRs. While pre-refereed drafts were slightly more often self-archived than other types of research content, the frequency of IR contribution is low in general.

The findings indicate that for the majority of self-archivers self-archived research papers or book chapters 5 times or less, which means that self-archiving is an occasional activity for faculty. Among the self-archiving Web venues, however, personal web pages were most frequently used for self-archiving, whereas IRs were least popular venue.

The present study received survey answers from 269 respondents whose materials were already deposited into their university IRs, although only 98 out of those 269 (36.4%) mentioned that they actually contributed to IRs. This suggests that most IR contributors did not realize that their materials were placed in IRs, probably since the submission to IRs was conducted by IR or departmental staff members rather than by the faculty members themselves. Also, 11 respondents from the IR non-contributor sample reported that they contributed to IRs. Thus, the total number of real IR contributors (those who identified themselves as IR contributors) was found to be 109 (15.9%). The following sections will examine survey respondents' awareness of DSpace IRs, as well as characteristics of the real IR contributors. It is important to examine these real IR contributors because they chose IRs by themselves as a venue for self-archiving. This personal choice indicated their motivation of IR contribution, which the present study attempts to investigate.

IR Awareness

The present study examined how many respondents were aware of IRs in their universities. Table 4.9 shows that 274 (40.1%) had some awareness of IRs, whereas 400 (58.5%) were not aware of the IRs in their universities. The questionnaire also provided

a link to DSpace IRs in each university in order to have respondents consider their university IR when answering the question.

Table 4.9. Awareness of IRs

	Freq.	*Percent*
Aware of IR	274	40.1
Not aware of IR	400	58.5
N/A	10	1.5
Total	684	100.0

Out of the 274 respondents who were aware of IRs, 193 (70.4%) were from the IR contributor sample. However, out of the 400 respondents who had no awareness of IRs, 330 (82.5%) consisted of those in the IR non-contributor sample randomly drawn from professors in twelve disciplines. The remaining 70 respondents were from the IR contributor sample, yet they were not aware of IRs probably due to the fact noted above that IR or library staff submitted their materials into the IR unbeknownst to them.

In addition, the 274 respondents who were aware of their university's IR learned about the repository in various ways. Table 4.10 shows that the majority of the respondents were contacted by IR staff. Other respondents learned about the IR through publicity about the IR via university/library web sites, word-of-mouth from faculty colleagues, and IR staff's presentations in a meeting. Overall, either direct contact or presentations made by IR staff was a primary method that had respondents aware of the IR.

Table 4.10. How respondents learned about the IR

	Freq.	*Percent*
IR staff contacted me	81	29.6
Publicity about the IR in university/library web sites	61	22.3
Faculty colleagues told me about the IR	53	19.3
IR staff made a presentation at meetings	46	16.8
Publicity about the IR in campus newspapers	36	13.1
Dean or chair of my school told faculty about the IR	31	11.3
Another university authority told me about the IR	28	10.2
Publicity about the IR via e-mail, mail or flyers from the library	7	2.6
Other	31	11.3

Thirty-one (11.3%) respondents also mentioned several other ways in which they learned about the IR. Nine reported that my survey helped them become informed of the IR, although these were not valid responses because the question asked whether they had awareness of the IR prior to answering the survey. In addition, two respondents served on a faculty advisory board for the IR, and two others became aware of the IR through a library committee regarding the repository. Moreover, two respondents found the IR through an Internet search, and two others mentioned that they accidentally found that their materials were deposited into the IR and thus, learned about it. One respondent also stated that technical reports in his department were automatically placed in the IR. Proxy submission, employed by several IRs, was a main reason that faculty awareness of the IR was not very high even if their materials were deposited there.

Real IR Contributors

As stated earlier, 109 (15.9%) respondents actually have contributed to the IR in their university. Table 4.11 shows the rank of 109 IR contributors from two sampled groups. Only 17 (15.6%) are assistant professors, whereas the rest of the IR contributors are tenured professors. In particular, 51 (46.8%) and 7 (6.4%) are full professors and distinguished professors, respectively.

Table 4.11. Real IR Contributors by Rank

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Assistant Professor	14	14.3	3	27.3	17	**15.6**
Associate Professor	32	32.7	2	18.2	34	**31.2**
Full Professor	47	48.0	4	36.4	51	**46.8**
Distinguished Professor	5	5.1	2	18.2	7	**6.4**
Total	98	100.0	11	100.0	109	**100.0**

As seen in Table 4.12, IR contributors are overwhelmingly Engineering 42 (38.5%) and Social Sciences 38 (34.9%) professors. Fewer professors from the Sciences contribute, although researchers in Physics and Mathematics are known to be actively

involved in self-archiving through disciplinary repositories such as arXiv.org, rather than IRs.

Table 4.12. Real IR Contributors by Discipline

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Science	15	15.3	1	9.1	16	**14.7**
Engineering	40	40.8	2	18.2	42	**38.5**
Social Science	34	34.7	4	36.4	38	**34.9**
Humanities	9	9.2	4	36.4	13	**11.9**
Total	98	100.0	11	100.0	109	**100.0**

Table 4.13 presents the distribution of real IR contributors based on their universities. **A** University has the largest number of collection titles in the IR and the number of its IR contributors is accordingly the greatest. Except for this case, the size of IR collection does not seem to be correlated with the number of real IR contributors. **B** University contains the second largest collections, whereas only five faculty members identified themselves as IR contributors. In **D** University where 7 IR contributors were identified (See Table 4.7), no respondents reported that they made contribution to the IR. It is common across the universities that merely a limited number of professors have contributed to their university's IR. This finding was consistent with results from other studies, which suggest that some operational IRs do not focus on faculty output, but rather on theses and dissertations or digitized archival collections (Markey et al., 2007; Yakel et al., 2008).

Table 4.13. Real IR Contributors by University

	IR contributors		**IR non-contributors**		**Total**	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
A	20	20.4	0	0.0	20	**18.3**
O	15	15.3	2	18.2	17	**15.6**
E	13	13.3	0	0.0	13	**11.9**
I	12	12.2	0	0.0	12	**11.0**
M	8	8.2	1	9.1	9	**8.3**
B	5	5.1	1	9.1	6	**5.5**
K	6	6.1	0	0.0	6	**5.5**
G	4	4.1	1	9.1	5	**4.6**

H	4	4.1	1	9.1	5	**4.6**
J	4	4.1	0	0.0	4	**3.7**
P	4	4.1	0	0.0	4	**3.7**
Q	1	1.0	3	27.3	4	**3.7**
C	0	0.0	2	18.2	2	**1.8**
F	1	1.0	0	0.0	1	**0.9**
L	1	1.0	0	0.0	1	**0.9**
D	0	0.0	0	0.0	0	**0.0**
N	0	0.0	0	0.0	0	**0.0**
Total	98	100.0	11	100.0	109	100.0

Self-archiving in the Future

The present study explored respondents' intention to self-archive in the future. As seen in Table 4.14, 474 (69.3%) plan to self-archive later, whereas 46 (6.7%) have no intention to do so. Furthermore, 149 (21.8%) are uncertain whether they will self-archive in the future. Out of 474 potential self-archivers, 416 (87.8%) are currently involved in self-archiving. However, 38 out of 46 (82.6%) respondents who do not self-archive also have no plan to self-archive. In addition, 100 out of 149 (67.1%) who do not self-archive are uncertain whether they would ever self-archive.

Table 4.14. Current Self-archiving vs. Future self-archiving

	Current Self-archiving					
Future Self-archiving	Yes		No		Total	
	Freq.	*Percent*	*Freq.*	*Percent*	*Freq.*	*Percent*
Yes	416	60.8	58	8.5	474	69.3
No	8	1.2	38	5.6	46	6.7
Uncertain	49	7.2	100	14.6	149	21.8
N/A	7	1.0	8	1.2	15	2.2
Total	480	70.2	204	29.8	684	100.0

In fact, current self-archiving was found to be significantly related to future self-archiving based on the result of the Chi-square test ($\chi 2 = 232.59$, $p=0.000$). This finding indicates that those who currently self-archived are more likely to participate in self-archiving practices some time down the road. Those who do not currently self-archive are less likely to plan to self-archive, or are unsure of the participation in future self-archiving.

Respondents who planned to self-archive reported in which venues they would be willing to do so. Table 4.15 shows that most selected personal web pages as a preferred future self-archiving venue, followed by departmental, and then research group web sites. Compared to those venues, fewer plan to self-archive in either disciplinary or institutional repositories. Other venues include MIT's Open Courseware site, PubMed Central, the PLoS, publishers' web sites, funding agency web sites, data publication series, such as that of the Ecological Society of America, public data archives, wikipedia or other wikis, open project supported by Mediawiki, Facebook or MySpace, open source repositories (e.g., simtk.org), Citeseer, and Social Science Research Network (SSRN).

Table 4.15. Self-archiving Venues in the Future

	Freq.	*Percent*
Personal web pages	384	81.0
Department/school/college web sites	276	58.2
Research group/lab/center web sites	265	55.9
Disciplinary repositories	174	36.7
IRs	159	33.5
Other	19	4.0

In particular, the present study examined what features of the IR would motivate respondents to contribute to the IR. Table 4.16 presents that long-term preservation of their work in the IR would be the primary factor, followed by the IR providing usage statistics for contributors.

Table 4.16. Motivating factors affecting IR contribution in the Future

	N	*Mean*	*Std. Dev.*
The ability of the IR to preserve my materials	228	3.94	1.15
The IR would show how many times the materials I deposited in the IR were viewed and downloaded.	224	3.40	1.17
Publishers would not have exclusive rights over my work.	225	3.17	1.32
Formal university recognition	221	3.16	1.29
The establishment of a peer review process in the IR	216	3.13	1.30
IR contributions would count toward my tenure and promotion.	236	2.41	1.34
I would receive financial reward.	223	2.09	1.11

The usage statistics are related to notability of research works, one of criteria for promotion. However, respondents do not seem to see the relationship between the IR contribution and tenure/promotion. Financial reward is the least motivating factor that would influence the IR contribution.

Factors affecting Self-archiving

The primary research question of this study is "what factors influence whether or not, and how much faculty members are involved in self-archiving." The previous section has discussed the question of how much. I will now turn to the question of the factors that influence self-archiving. Based on the literature review, I have identified eleven factors that affect self-archiving behavior. In order to measure those factors, 33 Likert-scale questions were developed and provided in the online questionnaire. Each factor was presupposed to be a group of two, three, or four scale questions. By calculating Cronbach's alphas, the reliability of each construct was determined (Table 4.17). During the process, three scale questions with low alpha values were excluded from subsequent calculations, and thus 30 scale questions were eventually used (See Appendix D).

Table 4.17. Reliability of Factors

Factors/Independent variables		*Number of Questions*	*Cronbach's Alpha*
Cost factors	Copyright concerns	3	0.71
	Additional time and effort	3	0.79
Extrinsic benefit factors	Accessibility	2	0.67
	Publicity	3	0.86
	Trustworthiness	3	0.74
	Professional recognition	2	0.83
	Academic reward	2	0.89
Intrinsic benefit factors	Altruism	4	0.80
Contextual factors	Trust	3	0.74
	Self-archiving culture	2	0.66
	Influence of external actors	3	0.67

As seen in Table 4.17, all but three factors show alphas greater than .70, the accepted level of internal consistency for items in social science research. The factors of accessibility, self-archiving culture, and the influence of external actors have alphas of .66 or .67, which is in the minimally acceptable level of reliability (DeVellis, 2003, p.95).

Individual scores of items incorporating each factor were summed up so that the sums of scores represented the constructs. Table 4.18 presents descriptive statistics for the summed scores, including means, standard deviations, minimum, and maximum values.

Table 4.18. Descriptive Statistics of Total Scores by Factors

Factors/Independent variables		*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
Cost factors	Copyright concerns	622	8.06	3.41	1	15
	Additional time and effort	662	8.00	2.91	1	15
Extrinsic benefit factors	Accessibility	660	7.94	1.86	1	10
	Publicity	665	12.02	2.34	2	15
	Trustworthiness	651	10.93	2.20	1	15
	Professional recognition	651	7.10	2.02	2	10
	Academic reward	591	7.12	2.25	1	10
Intrinsic benefit factors	Altruism	654	15.03	3.42	2	20
Contextual factors	Trust	631	8.31	2.83	1	15
	Self-archiving culture	635	5.80	2.19	1	10
	Influence of external actors	635	7.13	2.30	1	13

In addition, three variables relating to individual characteristics were included: (1) the number of publications per year; (2) rank; (3) technical skills. The publication number was derived from a multiple-choice question with four options: 0-1, 2-3, 4-5, and greater than 5. Those options were coded as 1, 2, 3, and 4, respectively. The rank denoted four professional titles including assistant, associate, full, and distinguished professor. These titles were also coded as 1, 2, 3, and 4, correspondingly. While those two variables were categorical, the variable measuring technical skills was continuous.

As seen in Table 4.19, respondents self-rated their technical skills on 1-5 scales. One signified no knowledge and five indicated expertise. The scores for each individual respondent were averaged. The average score represented the variable of technical skills.

Table 4.19. Self-reported Technical Skills

Technical skills	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
Send and receive e-mail	656	4.75	0.52	1	5
Search for information on the Internet/World Wide Web	655	4.53	0.67	1	5
Use a spreadsheet or database program on a computer	655	3.75	1.18	1	5
Create or edit a World Wide Web site (using such programs as html)	655	3.08	1.43	1	5
Program a computer using a programming language (such as C, C++, Java)	655	2.58	1.56	1	5

Other than those independent variables, the present study employed two dependent variables, each of which represented different aspects of self-archiving behavior. One dependent variable indicated percentages of five types of research materials – (1) pre-refereed drafts; (2) refereed articles; (3) unrefereed articles; (4) book chapters; (5) data sets - that respondents had self-archived in the past five years. The question measuring the variable provided five choices to select – none, 1-25%, 26-50%, 51-75%, and 76-100%. Individual scores of the percentages were summed up and the total score represented the dependent variable. Table 4.20 indicates the mean, standard deviation, minimum, and maximum value of the sum.

Table 4.20. Descriptive Statistics of the Continuous Dependent Variable

Dependent variable name	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
Percentage of research work self-archived	680	9.53	4.87	4	25

The other dependent variable was binary – whether or not respondents had contributed to their university's IR. It was used to identify factors influencing IR contribution. Since this variable was dichotomous, logistic regression analysis was performed to explore significant factors.

The continuous dependent variable, however, signified to what extent faculty respondents had been currently involved in self-archiving. Because the variable was continuous, Ordinary Least Squares (OLS) regression analysis was conducted to investigate factors predicting the proportion of research materials self-archived by respondents.

In sum, Table 4.21 indicates two dependent variables: one was used for OLS regression analysis, and the other was for logistic regression analysis.

Table 4.21. Dependent Variables

	Dependent variables
OLS regression analysis	Percentages of self-archived research work
Logistic regression analysis	Whether or not respondents have contributed to IRs

Table 4.22 presents the independent variables used for both OLS and logistic regression analysis. Cost, benefit and contextual factors were measured based on the scale questions, whereas individual traits were collected as demographic information.

Table 4.22. Independent Variables used for the Regression Analyses

	Independent variables
Cost factors	Copyright concerns
	Additional time and effort
Benefit factors	Accessibility
	Publicity
	Trustworthiness
	Academic reward
	Professional recognition
	Altruism
Contextual factors	Trust
	Self-archiving culture
	Influence of external actors
Individual traits	Age
	Rank
	Number of publication
	Technical skills

Factors affecting the Extent to Self-archive

This study examined the relationship between several independent variables and a continuous dependent variable. For this purpose, OLS regression analysis was utilized.

Table 4.23. Results of OLS Regression

Factors/Independent variables		*Standardized coefficient*
Cost factors	Copyright concerns	**-0.15*** **
	Additional time and effort	**-0.09** **
Extrinsic benefit factors	Accessibility	0.01
	Publicity	0.03
	Trustworthiness	-0.03
	Professional recognition	0.05
	Academic reward	**0.09***
Intrinsic benefit factors	Altruism	**0.28*** **
Contextual factors	Trust	0.02
	Self-archiving culture	**0.27*** **
	Influence of external actors	0.04
Individual traits	Age	**-0.10***
	Rank	**0.12***
	Technical skills	**0.14*** **
	Number of publications per year	0.03
	R square	0.55
	F	**36.98*** **

*p<.05; **p<.01; ***p<.001

Table 4.23 shows the results of the regression analysis. The R-square of .55 ($F=36.98$, $p<0.001$) means that 55% of the variation in the dependent variable can be accounted for by all the independent variables. The standardized coefficients indicate the extent to which positive or negative associations exist between each independent and dependent variable. In particular, six factors are found to be significantly related to the dependent variable. Two cost factors, including copyright concerns and additional time and effort, are negatively related, whereas academic reward, altruism, self-archiving culture, and technical skills are positively associated. The standardized coefficients of

those significant factors indicate the amount of increase or decrease in the percentage of self-archived research work predicted by a one unit increase in the factors. For example, the coefficient of copyright concerns is -0.15, which means that for every unit increase in copyright concerns, a 0.15 decrease in percentages of self-archived research work is predicted. In contrast, for every unit increase in altruism, there is a 0.27 increase in percentages of self-archived research work.

Copyright concerns and additional time and effort were proposed to be factors that would impede self-archiving, and the regression analysis showed these impacts. One respondent (ID 1052) specifically mentioned that "The major constraint on posting materials [is] copyright restrictions. First, some publishers make it difficult. Second, I sometimes use copyrighted material (e.g. tables and figures) in my lectures and it is a pain in the neck to get copyright clearance to post my lecture notes for public utilization." It was interesting to note that concerns about copyright related not only to respondents' own copyrighted materials, but also to other researchers' copyrighted works used in teaching materials. Although the regression analysis determined the relationship between copyright concerns and self-archiving research papers, books or data sets, copyright issues could involve various other types of materials created by faculty respondents, such as lecture notes or slides.

Additional time and effort was another factor that hindered self-archiving. One respondent (ID 2286) commented "The main reason I have not posted more items on my own (University-related) personal web page is just that I have not had time to do this, given all the other things I am expected to do". Similarly, another respondent (ID 1472) mentioned that time was a main reason for not posting her work on the Web and she ended her comment saying that "if someone else offered to do it for me, I would gladly accept!" The fact that faculty members did not consider self-archiving a primary task to accomplish, and it required spending an extra amount of time and effort, would lower the participation in self-archiving.

The regression analysis also determined four factors that would motivate self-archiving. One factor, academic reward, measured the degree to which respondents agreed with the statement that self-archiving would have no negative influence on tenure and promotion. The standardized coefficient was 0.09, which indicates that for every unit increase in the agreement of no harmful impact of self-archiving on tenure and promotion, there was a 0.09 increase in percentages of self-archived research work. In other words, respondents who perceived self-archiving to be unhelpful for their tenure and promotion, were less likely to post their research work on publicly accessible web sites. Since academic reward systems are deeply rooted in the peer review process, respondents seemed to suspect that there would be little positive effect of self-archiving on tenure and promotion, especially posting non-peer-reviewed materials. One respondent (ID 1433) noted "While peer review certainly has its problems, there is a little reassurance that the research quality is maintained. Also, I don't see how any of the self-archived documents could ever count toward tenure without documentation of rigorous peer review."

The factor of altruism measured the extent of agreement with ideas that self-archiving would help other researchers build on research, and allow them to have open access to respondents' materials that they could not use otherwise. The more strongly respondents believed these beliefs, the greater proportion of research work they had self-archived. In the sense that academic authors were interested in making their research easily accessible, self-archiving could be driven by the mix of altruism and self-motivation for increase in accessibility of research work.

Self-archiving culture was measured by the degree to which respondents agreed that self-archiving was common in their field or department. Respondents who agreed more strongly with the existence of self-archiving culture in their disciplines had self-archived a higher percentage of research papers, books, or data sets. In particular, self-archiving culture was closely related with disciplines respondents belonged to. Table

4.24 presents averaged scores on the factor of self-archiving culture by discipline. Science included Physics, Mathematics, and Biology and Engineering encompassed Electronic and Computer Engineering, Computer Science, and Mechanical Engineering. The Social Sciences respondents were from Economics, Sociology, and Psychology and the majority of Humanities respondents were in the fields of English, History, and Art.

Table 4.24. Average Scores on Self-archiving Culture by Discipline

Self-archiving culture by discipline	*N*	*Mean*[a]	*Std. Dev.*
Science	152	3.08	1.04
Engineering	179	3.22	1.01
Social Science	189	2.96	1.07
Humanities	115	2.06	0.92

[a] The means ranges from 1= Strongly disagree to 5=Strongly Agree

As seen in Table 4.24, respondents in Engineering agreed most strongly that self-archiving culture existed in their discipline or department. Humanities respondents, however, rated the existence of a self-archiving culture in their discipline very low. I conducted one-way ANOVA test to see whether the mean of scores on self-archiving culture differed among the four disciplinary areas. The result of the ANOVA test indicated that scores on self-archiving culture were significantly different among the four areas ($F=33.56$, $p<.001$). Specifically, scores on self-archiving culture in Humanities differed significantly from those in Science, Engineering and Social Science. However, there was no significant difference in scores on self-archiving culture between Science and Engineering, Science and Social Science, and Engineering and Social Science. Thus, self-archiving culture in Humanities was much weaker and significantly different than that in other disciplines.

Three individual traits – technical skills, age, and rank – were found to be significantly related to percentages of self-archived research work. Technical skills and rank were positively related, although age was negatively associated.

Concerning technical skills, the greater a respondent's rated themselves vis-à-vis technological expertise, e.g., relating to computers and the Internet, the higher the

percentage of research papers, books, or data sets they had self-archived. This factor also is also associated with respondents' disciplines, and in particular, respondents from Engineering reported much higher scores on technical skills than those in other disciplines (See Table 4.25). I performed one-way ANOVA test which resulted in a significant difference in the means for technical skills between four disciplinary areas ($F=48.33$, $p<0.001$).

Table 4.25. Average Scores on Technical Skills by Discipline

Technical skills by discipline	*N*	*Mean*[a]	*Std. Dev.*
Science	154	3.79	0.79
Engineering	184	4.16	0.73
Social Science	196	3.65	0.72
Humanities	118	3.14	0.67

[a] The means ranges from 1= No knowledge to 5=Expertise

The professional rank of respondents was another individual trait positively related with percentages of self-archived research works. The rank was coded assistant, associate, full, and distinguished professor as 1, 2, 3, and 4, respectively. Therefore, tenured professors tended to self-archive a greater percentage of research materials than untenured professors.

Finally, age was found to be negatively associated with the dependent variable, which suggested that younger faculty members self-archived a greater percentage of research works than older professors. One respondent stated that "I have been a professor for over forty years and am neither as attracted to or skilled at using the web as a vehicle for disseminating research as my younger colleagues" (ID 1205). The correlation between age and technical skills was also typically assumed. The further discussion regarding age and technical skills was made in the section of interview data analysis.

Factors affecting IR Contribution

An IR was the most recently introduced venue for self-archiving. As noted, I found 109 (15.9%) respondents who actually had contributed to their university's IR. In

the present study, I was interested in understanding the factors that influenced whether or not respondents deposited their research work in IRs. Because this dependent variable was dichotomous, it was appropriate to conduct logistic regression analysis.

Table 4.26 shows the results of logistic regression analysis, which identified five significant factors affecting IR contribution: (1) copyright concerns; (2) accessibility; (3) altruism/self-interest; (4) trust.

Table 4.26. Results of Logistic Regression Analysis regarding IR Contribution

Factors/Independent variables		*Odds Ratio*	*z*
Cost factors	Copyright concerns	**1.10***	2.12
	Additional time and effort	0.96	-0.86
Extrinsic benefit factors	Accessibility	**1.29***	2.30
	Publicity	1.11	1.05
	Trustworthiness	0.92	-1.29
	Professional recognition	0.86	-1.63
	Academic reward	0.98	-0.25
Intrinsic benefit factors	Altruism	**1.18****	2.64
Contextual factors	Trust	**1.15***	2.35
	Self-archiving culture	0.90	-1.43
	Influence of external actors	1.07	1.13
Individual traits	Age	1.01	0.28
	Rank	1.48	1.86
	Technical skills	1.20	0.93
	Number of publications per year	0.94	-0.36

*p<.05; **p<.01

Unlike OLS regression analysis, logistic regression generates odds ratios, which indicate the amount of change expected in the likelihood of IR contribution when there is one unit change in an independent variable with all of the other variables held constant. For instance, the odds ratio of copyright concerns is 1.10, which means that the odds of IR contribution increase by a factor of 1.10 for every one unit change in copyright concerns when all other variables are held constant. This finding suggests that respondents who were more cautious about copyright issues were more likely to

contribute to their university's IR. This was an opposite result to the effect of copyright concerns on percentages of general self-archived research work. Nevertheless, IRs were implemented by university libraries committed to provide services complying with copyright law, and potentially this motivated respondents concerned about copyright issues to participate in IR contribution.

The factor of accessibility was measured based on the extent to which respondents agreed that self-archiving would increase the chance of communicating their research findings with peers, and make their research work more easily accessible through search engines. Respondents who more strongly agreed with the idea of accessibility of self-archived materials were more likely to contribute to their university's IR. Functionalities of IRs that facilitated accessibility, such as permanent URIs and OAI compliance, might encourage respondents to use the IR as one of their self-archiving venues.

Altruism was also found to be a significant factor affecting IR contribution. As seen in Table 4.26, the odds of IR contribution increased by a factor of 1.18 when every unit in altruism increases with all other variables held constant. Respondents with a higher degree of reported altruism or self-interest in open access were more inclined to deposit their materials into the IR.

Trust was measured in several ways: the degree to which respondents agreed that the good intentions of readers using their self-archived materials appropriately, their perceptions of competence of readers, and their perceived ability of institutions to perform secure maintenance of materials and correctly manage self-archiving venues. The logistic regression analysis shows that respondents with more trust in readers and institutions were more likely to contribute to the IR.

The OLS regression analysis determined eight factors significantly related to percentages of self-archived research work, which represented the extent to which faculty respondents participated in self-archiving activities. The logistic regression analysis identified four factors associated with faculty IR contribution. The significant factors

found in the analyses were further explored based on the analysis of interview data in the following sections. The interview data analysis also introduced new constructs, which were not examined in the survey.

Summary of Survey Data Analysis

Out of 684 survey respondents, 480 (70.2%) have made their research publicly accessible on the Web. Although this large proportion of survey respondents has self-archived, the extent that they are involved in this practice varies. The largest number of faculty self-archivers has made refereed articles publicly accessible. Around 68% of those who self-archived peer-reviewed papers posted more than half of their refereed articles on publicly available sites (Figure 4.1). Refereed articles are the most typical genre for self-archiving, although a variety of other types of research materials, e.g., data sets or presentation slides, have been self-archived (Figure 4.2).

Personal web pages were used by the majority of self-archivers and utilized more frequently than other venues (Figure 4.4). IRs were the least popular venue for self-archiving, although this study identified 109 real IR contributors who I regard as early adopters of IRs. The study hypothesized that their perceptions of self-archiving would differ from those of non-IR contributors. The logistic regression analysis was performed to test this general hypothesis and to determine factors affecting whether or not faculty would contribute to IRs (Table 4.26). As a result, four factors were found to be positively associated with IR contribution in particular: (1) copyright concerns; (2) accessibility; (3) trust; (4) altruism.

This study also conducted the OLS regression analysis to examine the relationship between various factors and the percentage of self-archived research to measure the extent to which faculty have self-archived. This analysis was important to answer my core research question – which factors motivate or impede self-archiving. Overall, five

factors were positively associated with self-archiving: (1) altruism; (2) academic rewards; (3) self-archiving culture; (4) technical skills; (5) rank. Three factors were negatively related to self-archiving: (1) copyright concerns; (2) additional time and effort; (3) age.

Interview Data Analysis

The series of regression analysis identified several factors that influence faculty respondents' decision to self-archive. These factors were explored in greater depth in 41 follow-up phone interviews. Thirty-two interviewees were self-archivers, whereas nine were non-self-archivers. Table 4.27 and 4.28 present the distribution of interviewees including self-archivers and non-self-archives based on their academic rank and disciplines, respectively. Self-archivers were recruited from 12 universities, whereas non-self-archivers were from 7 universities.

Table 4.27. Interviewees by Rank

	Self-archiver	*Non-Self-archiver*
Assistant professor	13	5
Associate professor	5	3
Full professor	14	1
Total	32	9

Table 4.28. Interviewees by Discipline

	Self-archiver	*Non-Self-archiver*
Science [a]	14	2
Engineering [b]	6	5
Social Science [c]	8	1
Humanities [d]	4	1
Total	32	9

[a]. Mathematics, Molecular Biology, and Physics
[b]. Computer Science, Mechanical Engineering, and Electrical and Computer Engineering
[c]. Sociology, Psychology, and Economics
[d]. Art, English, History, Linguistics, and Slavic Languages

The analysis of my interview data uncovered a variety of experiences and opinions about self-archiving behavior and helps to explain some of the findings from the survey data. The interviews confirm motivating and impeding factors identified in the survey data, and convey some aspects of self-archiving behavior that my survey did not attempt to quantify. The following sections describe the multiple types of content and venues that interviewees referred to for self-archiving, and the reasons that they make or do not make their research publicly accessible on the Web. I then discuss a wide range of factors that might influence professors' decisions to self-archive research and/or teaching materials.

Interviewees discussed a variety of content types that they had made publicly accessible on the Web and explained reasons for their self-archiving. They also mentioned content types that they preferred to keep private and in what circumstances they would do so. The decision of what to make publicly accessible and what not to depends upon the judgment on the quality of the work, disciplinary norms, and copyright issues.

Research Materials

The present study defines self-archiving as making research materials publicly accessible on the Web. Research-related materials that interviewees self-archived range from pre-refereed, unrefereed, and refereed articles to data sets, lab protocols, posters, and software programs. When discussing self-archiving refereed articles, ten interviewees stressed the value of the peer review process as a quality control mechanism. By making refereed publications publicly available, interviewees noted that they were able to claim an idea was theirs and had no worries about getting scooped when self-archiving. In addition to the journal peer review process, two physicists addressed the importance of the internal peer review process within their large scientific collaboration.

Peer Review Process

Except for professors in Physics or Mathematics who used arXiv, four interviewees in other disciplines were unwilling to post papers, which had not gone through a peer-review process or been accepted or published by a journal, on a publicly available web site. Those professors across various disciplines – Electronic and Computer Engineering, Computer Science, Sociology, and Psychology - commonly believed that their papers without peer review would be inadequate so that they were not ready to post them publicly on the Internet. In particular, a full professor in Electrical and Computer Engineering mentioned:

> I prefer to work with posting stuff that has gone through peer review just so that I know that it's been checked by others of an anonymous nature who aren't afraid to tell me because of the anonymity that oops you goofed here or that's not quite right (ID 2090).

Posting research published in peer-reviewed journals was seen as a way to protect their ideas from being stolen by other researchers. The professor in Electrical and Computer Engineering was reluctant posting research work that "haven't been vetted for correctness of ideas" (ID 2090). If he made it publicly accessible on the Web, others might have a chance to explore it before he did. Similarly, an associate professor in Mechanical Engineering who has not self-archived any of her work (ID 1919) stated that she did not post papers sooner than they were published so that she did not get scooped on her publications.

Another reason not to post research work under review was that it was necessary to remain anonymous while the work was undergoing peer review. An associate professor in Information Retrieval (ID 1499) mentioned that conference papers in his field went through a blind review and thus he did not make those under review publicly available on the Web. An assistant professor in Economics (ID 2251) also pointed out that he kept papers under review private because he desired to remain anonymous.

Other than the peer review process for journal publications, an internal review process in large collaboration projects functioned as a quality assurance mechanism. Two physicists stated that their collaborators cross-checked the quality of research conducted through the collaboration and decided what results to make publicly accessible on the Web. One associate professor in Physics mentioned that he and his collaborators had vast amount of data in experiments, which it would take a very long time to fully analyze. Their work process is such that they usually make several "interim updates" (ID 2031) before publishing a final version which consists of all the updates. Although each update contains only partial and preliminary results, they felt that the students who worked very hard on the small parts of the analyses needed the opportunity to present their work in public prior to the final report. The professor also suggested that:

> It is important to understand that in the case of these large collaborations any result that comes out is carefully reviewed internally by these committees that are set up. One of the advantages of having so many people on the author list is it's possible to find people who are not actually working on the analysis to do the review (ID 2031).

The other associate professor in Physics working in a large experimental collaboration emphasized the responsibility of maintaining scientific accuracy in the fundamental research that he and his collaborators conducted while still making results publicly accessible on the Web. He stated that:

> My credibility and the credibility of my collaborators is on the line and what we release are finished analysis results that we have all agreed and done cross checks on and thought about at great length and we carefully release those results and papers that interpret them…it's more in the interest of credibility and scientific accuracy rather than preserving something that you're going to make money out of (ID 1925).

Versions of Refereed Articles

Although publishers often did not permit authors to self-archive publishers' PDF versions of refereed articles, interviewees preferred to self-archive them. An assistant professor in Mechanical Engineering (ID 1435) would not post her article "that's in

print" unless publishers permitted her to do so. If she did post articles, the publisher's PDF version would be her first choice to make publicly available on the Web. Instead, she self-archived earlier versions of her published article.

However, three professors made publishers' PDFs of refereed articles publicly available regardless of copyright restrictions. They preferred publishers' PDFs since it was the final version that colleagues were reading in the actual journal. Specifically, professors in Mathematics and Physics who submitted papers to arXiv and then published those later in journals self-archived the publisher's PDF version once it was published. The assistant professor in Mathematics (ID 1916) described that authors were not so great about correcting their papers based on referees' comments that papers on the Web were not a final form in most cases, whereas published ones were a truly final version. He believed that if people could access both the published version and arXiv version of the same paper, they would certainly read the published one. Publishers did not seem to care that authors self-archived PDF versions on the Web because a number of people reading these papers at institutions had subscriptions to all the journals. A full professor in Physics (ID 2337) even noted that it was probably "illegal" to take publishers' PDFs from journals to post them on his web page but he did not ask publishers whether it was allowable.

The assistant professor in Ecology also mentioned publishers' PDFs were much more readable than her own final version, which was "like typeset when the figures are not interspersed into the document" (ID 1505). She posted all publishers' PDFs of her refereed articles on the Web because she knew a couple of publishers allowing authors to self-archive publishers' PDFs and assumed other publishers might do so.

Those interviewees who self-archived publishers' PDFs emphasized the difference between earlier versions of papers and the published versions in terms of either content or format. One assistant professor in Sociology who worked at a research center focusing on disaster issues stated that she usually posted preliminary papers on the Web

because some of the findings relevant to a specific event needed to be public soon. In the mean time, she also published articles revised and polished from the preliminary papers. She however, expressed concerns about the possibility that readers would use only preliminary papers, which was different from published articles because the earlier papers were easier to access through the Internet than journal articles, as described by the interviewee:

> I have to worry about if someone is interested in a particular paper, a particular topic, that they're just going to go to the preliminary paper and they may never go to the actual final draft that may have nuances or changes that maybe very different than that preliminary version. So it's really making that judgment call that the preliminary paper is complete enough, that I'm okay that might be the only source that someone look at (ID 1702).

This interviewee also noted that she did not have permission to post the published version on the Web. As a result, her preliminary papers were "more often cited than the article that's been published". As she described it:

> Really when the published article theoretically and more from an academic perspective seems much more relevant to some of the work in which the first paper is being cited so the fact that people can easily access it online rather than paying for the journal or having it through inter-library loan they'll opt for that (ID 1702).

Concerning whether it was legal or not to post publishers' PDFs freely on the Web, one full professor in languages and linguistics discussed his experience of consulting a major expert on copyright. Since the professor was an editor of a journal, he was interested in establishing the journal policy to allow authors to make their articles available widely. Based on consultation with this expert, he realized that making publishers' PDFs available on the Web would not cause serious legal problems:

> As far as I can tell there's not a clear and substantial problem with that and I speak as a journal editor there and also as someone who wants to see access to scientific materials as broadly as possible. So if somebody can tell me; if somebody can show me that there is a legal problem there I will absolutely stop doing that but I haven't seen that yet (ID 2214).

However, as an author, the professor usually self-archived his final version of refereed articles while he considered that there was little difference between the final version and a publisher's PDF version. He made a comment that it did not matter much "if it [the final version]'s the substance of the argument and the evidence that I'm most concerned with".

Data sets

Self-archiving data sets was a common practice in certain disciplines, especially Ecology and Molecular Biology. Interviewees who publicly shared data sets believed that the open data sharing enabled other researchers to build on research so that it would eventually advance science in an efficient way. A society journal or funding agencies, such as NSF also played an important role in facilitating the data sharing practice.

An assistant professor in Ecology (ID 1505) stated that many journals in her field required authors to deposit raw data into the National Center for Biotechnology Information (NCBI) databases, which are publicly available on the Web. Once depositing her data, she received a unique identifier for these data to be included in the paper when submitting to a journal. She mentioned that the publicly accessible data sets were commonly used by researchers in her field.

The other assistant professor in Molecular Biology also explained that in addition to self-archiving published articles on the Web, he posted his raw data on TreeBase, a publicly available database for data sets used to make phylogenetic trees. By making the data publicly accessible, the professor pointed out that people would clearly understand how he had analyzed the data and thus it helped them replicate his work:

> By posting the data sets and the command files for the computer algorithm that I use to make the trees and posting the trees themselves then that really allows other users to see all the nitty gritty of exactly what I did and what steps I took and all the algorithms that I ran in order to produce the results that I got because you really just can't do that in the paper itself (ID 2115).

A full professor in Ecology mentioned registering metadata for raw data, which the Ecological Society of America (ESA) strongly encouraged. When authors submitted a manuscript for publication to one of the Society's journals, they were also urged to deposit metadata in the data registry program. To do this, authors submitted "data papers", which meant "the combination of the metadata and the data themselves… not only the raw numbers but an explanation of how they were collected, how they were measured, how they were being reported" (ID 1507).

The professor (ID 1507) specified two incentives that motivated him to self-archive data sets. First, he had such a large amount of data that he could not publish findings from all aspects of it in his life time. He specified that "the science is really not of value to other people unless it's accessible". Self-archiving made it possible to publicize his data effectively since numerous people around the world had access to the Internet. Second, the National Science Foundation (NSF) strongly suggests that those who receive public funding make their data collected with grant funds publicly accessible on the Web. The two assistant professors in Ecology and in Molecular Biology also confirmed that NSF definitely encouraged them to self-archive data sets. Nevertheless, one of them commented that NSF was more concerned with publications than with the data set and "maybe it's by second relations that the data must be submitted in order to publish the papers and then NSF wants you to write papers or you cannot get any more funding" (ID 1505).

Like those in the Biological Sciences, one full professor in Electrical and Computer Engineering maintained his research group web site and provided huge amounts of data for scatterometer climate research sponsored by NASA. Besides this raw data, reports generated from the data sets and documentation regarding how to use the data were all publicly accessible. This self-archiving practice was driven from the contract with NASA to supply the data for them. He was also required to identify users of the data and report this back to NASA. The professor mentioned that he sometimes

encountered a problem when other scientists misused these data or incorrectly found a flaw in the data which resulted in an error in these scientists' analyses. Such difficulties occurred a couple of times per year, "often enough that there's a pain" (ID 2090). Whenever it happened, revised documentation and instructions about how to properly use the data were posted on the Web in order to remedy the problems.

In fact, the misuse of data was one of the concerns that made interviewees reluctant to self-archive data sets. A full professor in Developmental Sociology (ID 1803) described his concern that his data could be misused by those who did not thoroughly understand the context of his studies that usually focused on very specialized topics. Another assistant professor in Mechanical Engineering also stated that he worried about other people either claiming his data for themselves, or falsifying the data publicly available on the Web. However, he suggested that as long as he had the original data sets, he could cope with the misrepresentation problem. It would be a more serious problem, however, if he self-archived data sets including "something brand new and novel". He was afraid that someone might take that and publish journal articles and then he "wouldn't get recognition for having done that work" (ID 1976). One assistant professor in Economics expressed a similar concern that:

> There's more stuff that could be done with the data set. You've just done one thing and there's more that could be investigated and used. Particularly in my area, I do a lot of survey research so you spend a tremendous amount of time designing your instrument, collecting your data, and if you just put it up maybe other people will write papers (ID 1999).

The assistant professor added, though, that it would not be all that dreadful if she were cited by those who wrote papers based on her data.

Human subjects issues also made faculty reluctant to consider self-archiving raw data. An assistant professor in Health Informatics (ID 1958) noted his raw data dealing with a number of clinical charts and protected health information. Before posting these data on the Web, he made sure that they were completely de-identified. The assistant

professor was willing to self-archive the de-identified data if someone was interested in them. In contrast, a full professor in Psychology (ID 1457) stated that without identifying information, his data were not at all useful for researchers in his field, and therefore he kept them private.

Unrefereed Articles and Other Research-related Materials

While unrefereed articles were not commonly mentioned by interviewees for self-archiving, a couple of professors mentioned that self-archiving was the only way to make such articles accessible to the public. An associate professor in History noted that he wrote conference papers, which had "no venue to publish in written format" in his field, although he often cited other scholars' conference papers and was frustrated that he was unable to locate those papers either on the web or in written form. In this respect, he decided to make his conference papers publicly accessible on his personal web site. He also stated that one paper he wrote outside his subspecialty was not passed through peer-review processes in journals, and he finally decided "to circumvent the usual written publication format and put it up on the Web" (ID 1756). He told a story that one researcher in Jordan found the paper and asked him whether he could cite it. This experience reinforced his decision to self-archive papers that took a long time to be published or had no channel for publication because he thought that certainly there was a "demand" by other scholars who wished to read and cite the papers.

When considering technical reports or working papers, a full professor in Psychology had three or four reports that included important technical information, although it was "not archival quality for a journal" (ID 1457). Thus he just made them publicly accessible on the Web. On the contrary, a full professor in Computer Science never made working papers or white papers publicly available on the Web since the papers contained important ideas that he could use for future proposal. He wanted "the

ideas to remain in [his] ownership" (ID 1515) until he decided to write a proposal about them.

Other than traditional research papers, a range of research-related materials – software programs, presentations, posters, tables, figures, and lab protocols – were discussed by interviewees as objects for self-archiving. Three professors in Engineering have made software programs and related documentation publicly available on the Web. A full professor in Computer Science (ID 1984) posted programs without copyright protection on his personal web site, whereas another full professor in the same discipline (ID 1515) posted only copyrighted software and documentation on the Web. An associate professor in Mechanical Engineering (ID 2156) posted software and documentation developed under contract to the U.S. Department of Energy on its publicly available web site.

A couple of professors self-archived PowerPoint presentations. An assistant professor in Computer Science (ID 2238) pointed out that looking at his presentation slides before reading his paper made it easier to understand his research. Presentation slides were also useful for a specific audience. One assistant professor in Sociology mentioned that her PowerPoint presentations regarding disaster research had "a great deal of appeal for practitioners" (ID 1702) in emergency management positions. They were more likely to read presentations rather than journal articles attempted to be read by scholars.

In addition, an assistant professor in Health Informatics (ID 1958), a contributor to the IR at his university, commented that he wanted to preserve the content of his posters, such as diagrams, charts or tables in the IR. These "non-traditional academic outputs" were often used and displayed when he discussed his research with collaborators, and other people could access such materials. The professor also deposited supplementary materials – figures or tables – to published articles into the IR. Those materials were eliminated from the published articles according to a journal editor's

request, although he believed that the supplement was still useful for other scholars, as well as for him to reference it.

Laboratory protocols were self-archived by an assistant professor in Ecology (ID 1505). She put these on the Web as a resource for her students when she was not available, or to help researchers at other universities who investigating similar topics. However, she pointed out that not everyone thought posting lab protocols on the Web was good. Actually, one associate professor in Mechanical Engineering (ID 1919) asserted that she has never made her lab protocols publicly accessible on the Web. She shared them only with her collaborators or someone who "would acknowledge [her] expertise at the appropriate times". This professor tightly controlled her protocols on account of the fact that:

> The specifics of protocols….there are certain tricks of the trade in research that you don't always want to share all of those details with everyone because again that's what makes you more competitive for one grant or a research publication than another area. And so I think there's a certain amount of detail that we need to be protective of in order to stay in business (ID 1919).

In sum, faculty self-archivers preferred to post published, refereed articles on publicly accessible web sites because their quality was assured by peers and they did not have to worry about getting scooped. Publishers' PDF versions were the most preferable form for self-archiving, although many publishers restricted authors from doing this. Yet, interviewees did self-archive publishers' PDFs because they were finalized and readable. Self-archiving data sets was also common in Ecology and Molecular Biology and interviewees in these fields highly valued this practice. However, interviewees acknowledged the tremendous amount of effort required to manage these data sets and expressed concern about the misuse of data which made many interviewees reluctant to self-archive data sets. Interviewees also self-archived various other types of research works, such as presentations, posters, figures or tables. Because these materials were

usually not published, self-archiving was regarded as the only way to disseminate them widely.

Teaching Materials

The survey data analysis indicated that a majority of the faculty respondents made course syllabi and lecture notes publicly accessible on the Web (Table 4.2). Interviewees cited a number of reasons including response to students' needs, support for student learning, and no perceived negative impact by self-archiving teaching materials. One assistant professor in Economics mentioned that "there's a very strong student demand even at undergraduate and graduate level to make stuff publicly available and if I post something late or post the wrong thing they e-mail me" (ID 2251). Another full professor in Electrical and Computer Engineering stated that "my interest is more making this material freely available and easily accessible to my students to assist in their learning" (ID 2235). While he considered self-archiving teaching materials good for his students only, he unexpectedly received e-mails from someone in another country who appreciated his teaching materials, which the person used to develop his own course. The professor noted that this was truly an advantage of self-archiving, which he had never intended. The other reason for self-archiving teaching materials was described by one assistant professor in Computer Science that "I don't see why not and it's easily accessible to the students and if others are not enrolled in the class will look at my slides I don't see the harm in that" (ID 2238).

In addition to lecture notes or course syllabi, unpublished textbooks were self-archived by two other professors in Mathematics. One full professor posted a calculus book and a linear algebra book along with lecture notes of classes that he taught on his personal web page. He believed that his books were much better than other published textbooks, although students had to spend more than $100 purchasing the published ones.

He decided to make his books publicly accessible on the Web so that people who wanted to learn calculus or algebra would use the books for free. Another reason for self-archiving those books was that it was easy for him to revise the content if he found a mistake. He commented that "I just change it when it needs changing and that way I can make it a better book" (ID 2112). He reportedly invested substantial time writing and revising the books, but nevertheless he was pleased if anybody could use them freely from the Web. However, the other full professor who self-archived his unpublished book had a different reason to do so. He wrote the book 10-15 years ago as part of a monograph but he did not have enough time to make it publishable. Even though the book was "in a very unpolished form" (ID 2227), he made it publicly available on the Web so that anyone who liked it could have it for free. He did not want to "bother with publishing it because it would take a lot more work".

Interviewees also had concerns about copyright of content that they utilized in teaching materials. One assistant professor in Molecular Biology stated that he tried not to use copyrighted resources in his teaching materials but when he used those, he was very careful about attributing them appropriately. He posted his teaching materials on his university's WebCT so that only enrolled students could access them. One associate professor in History more strictly controlled his teaching materials due to copyright issues:

> I think that there are teaching materials that I keep private because to put them online would be possibly to violate copyright law. When I teach, I take images and other things from books that may not be able to put them on the Web without breaking copyright (ID 1756).

Similarly, another full professor in Mathematics explained that he often scanned pages of textbooks for teaching and put them on a password-protected web site. He noted that the scanned pages were "copyrighted and I don't want to break the law about putting other people's copyright material on the Web for free" (ID 2227).

The other interesting case, which the Mathematics professor discussed, was that he had not created teaching materials directly but used other researchers' teaching materials with their permission. He taught a graduate course with materials developed by a professor in Carnegie Mellon University and the materials were on a password-protected web site. The CMU professor gave him a username and a password, and the interviewee gave his students the access codes and asked them not to publicize it. The professor felt that this was in keeping with the original creator's intentions.

Other than copyright concerns and the effort involved in creating teaching materials, professors in the Humanities tended not to self-archive teaching materials because of their evolving nature. One full professor in Slavic Languages and Literatures stated that he constantly updated his teaching materials even throughout the semester. Since the materials were "always in progress" (ID 1467), he was reluctant presenting them on a publicly accessible web site. Another full professor in German Studies and Linguistics who was reluctant to post teaching materials noted that without understanding the full context for the materials, they could be easily misunderstood. Thus he handed out the materials only to his class, although he supported open access in principle.

Venues for Self-archiving

As found in the survey data analysis, personal web sites were also the most commonly utilized form of self-archiving by the interviewees across several disciplines. The large degree of control over material on personal web sites was regarded as a main reason for preferring this venue. From personal web sites, quite a few professors linked to their research group web sites, papers in arXiv, or course web sites so that people who accessed the personal web sites were able to see more resources. For that reason, the distinction between personal and research group websites was not very meaningful for some professors. These two web sites were often described as being "intertwined" (ID

2031) or "a conglomerate" (ID 1505). Yet, there were some professors who maintained solely research group web sites and self-archived their research work there. Departmental web sites were rarely used for self-archiving research work, but were sometimes adopted for posting newsletters or teaching materials.

In Physics and Mathematics, arXiv was an Open Access preprint server widely accepted by scholars. Most interviewees in these disciplines were heavy users of and contributors to arXiv. It was the norm to post papers on arXiv before submitting to a journal in these disciplines. Interviewees were also satisfied with the ease of self-archiving in arXiv, as well as benefits resulting from it. Compared to arXiv, Institutional Repositories (IRs) have a much shorter history and have not been used by a majority of scholars for self-archiving. Several interviewees, however, were aware that their universities had implemented IRs and expressed their opinions on the repositories. Seven IR contributors were also identified, and their experience of and attitude toward IR contribution were examined. The following sections investigated the interviewees' reasons of self-archiving on the following venues: (1) Personal web sites; (2) Research group web sites; (3) arXiv; (4) Institutional Repositories (IRs).

Personal Web Sites

The primary reason that interviewees self-archived on personal websites pertained to amount of control they could have over the web-based content. Seven interviewees perceived a great advantage in being able to post, revise or delete the content whenever they wanted. One assistant professor in Sociology described that:

> My personal web site is what I use more often because it's something that I have control over. I can update it and it's easy for me if I want to post something at 2:00 in the morning and that's when I want it done. I don't have to worry about someone else's schedule so the amount of control that I have, what's there, how it's presented is useful (ID 1702).

This control issue was frequently cited as a reason that interviewees would not adopt other types of self-archiving venue, especially the IR. One assistant professor in Ecology

felt more comfortable with her personal web site than the IR where someone else was in control. She commented that "I wouldn't want to put something in [the IR] and find out it's wrong or it needs to be changed and have them say well it cannot be removed and then I'm stuck" (ID 1505). In the same vein, an assistant professor in Molecular Biology (ID 2115) noted that he had not contributed to the IR since it was much easier for him to manipulate his own site than having IR personnel do it.

In addition, personal web sites were considered to be easily accessible and seen everywhere, particularly through Google. One full professor in Mathematics mentioned that "it's easy to get to because you can just Google [my name] and click on the top thing and there I am so it's easy to get" (ID 2112). A full professor in History (ID 2214) also stated that a researcher's personal website would be the first place that he and his colleagues looked for his/her papers. Likewise, a full professor in Physics (ID 2337) maintained his own web site because future graduate students or other people who might read his papers looked at it.

Faculty members made hyperlinks to their research works publicly available somewhere else, in addition to posting research/teaching work directly on personal web sites. Five interviewees provided a link to research group web sites, pre-refereed articles in arXiv, or course web sites. An assistant professor in Ecology (ID 1505) pointed out that she tried to place content concerning her research in one place, and thus other people, especially prospective students could see the breadth of her research area. Her personal web site consisted of some pages about her research, as well as several links to her current students' research. She mentioned that those were all associated, rather than separate from one another.

A couple of professors in Physics and in Mathematics made a link to the ArXiv version of their papers from personal web sites. Providing links to arXiv papers was a convenient way of maintaining personal web sites. Two other interviewees also retained a link to class web pages maintained by their departments on personal web sites.

Although the majority of interviewees utilized personal web sites for posting their work and advertising what they were doing, one associate professor in Physics wrote a blog, which provided information for his family and friends. He stated that it began as "an outreach program in physics when they were trying to emphasize people over science because in science we tend to talk too much about the science and not enough about the people" (ID 2031). The blog provided "a segued view of the outside world" so that he tried not to post too much about science in the blog.

Research Group Web Sites

Research group web sites were either separated from or connected with other venues. Sometimes, these web sites contained a mix of public and access restricted areas. For instance, one associate professor in Physics (ID 1925) who participated in a very large collaboration, used various research group web sites including one in his department server, a wiki situated on the servers of his collaborator, and password-protected servers at national laboratories. The local web site and the wiki also consisted of public and password-protected sections. In fact, the professor typically used a publicly accessible web space, which was not designed as a web page, to self-archive some materials. He let people know its URL if asked, and they could obtain the files there.

Research group web sites were also seen as beneficial: a full professor in Development Sociology (ID 1803) believed that maintaining his research group web site was very important because it allowed him to reach those with similar research interests. An associate professor in Physics (ID 2031) also stated that since his research group web site functioned as an archive of what he had done, it was extremely helpful that he could easily access it from anywhere when he was traveling.

Still, maintaining research web sites was considered problematic. An assistant professor in Sociology (ID 1702) explained that her research center had trouble hiring and keeping an IT person to maintain their web site. As a result, the web site was

outdated and she posted papers on her personal web site. She stated that without consistent IT support, it was difficult keeping the research web site up-to-date. Another assistant professor in Physics (ID 1839) suggested that often times other scholars' research groups web sites were out-of-date and thus were not useful very much.

arXiv

Self-archiving pre-refereed articles in arXiv was a disciplinary norm in Physics and Mathematics. Since arXiv was broadly adopted in these disciplines, virtually everyone reads papers in arXiv, as described by a full professor in Physics:

> If anybody wants to read my work then [arXiv]'s where they're going to go and if it's not there then they're not going to read my work and everybody else does the same thing at least in my field (ID 2452).

Most interviewees using arXiv were content to distribute their pre-refereed articles through arXiv for a variety of reasons. One advantage was that arXiv enables researchers to disseminate earlier versions of papers, instead of waiting for the published one in a journal. An assistant professor in Mathematics stated that since it took at least a year to publish his articles via a journal even after the peer-review process, it was more desirable to make earlier versions publicly available rather than waiting for the journal publication. Interestingly, he mentioned that early postings on arXiv provided him with a feeling of accomplishment and progress as explained that:

> I do see the benefit of early dissemination. For me it's just the personal one. For me I've accomplished – I've finished this paper – at least at the draft level. I've got a product available for people to view. I may have to make revisions later but at least I have a product done at that point – a preliminary product done (ID 2109).

The interviewee did not receive feedback about his early papers directly through arXiv. Rather, he has gotten it at conferences by talking to people. This was in contrast to what another full professor in Physics (ID 2337) mentioned. Before submitting a manuscript to a journal, the professor posted it on arXiv and waited two weeks or a month for feedback

from other people in his field, Theoretical High Energy Physics. He stated that there were around 1,000 very active scientists in the field, and he knew all of them, at least their names, who commented about his work on arXiv. Their comments were very useful when he revised the arXiv version for submission to a journal.

A couple of professors were fond of the interface of arXiv, viewing it as easy and simple to post articles. arXiv also stamps the date and time of posting on a paper and because of this function, interviewees were not worried about other people stealing ideas from their pre-refereed articles. Based on the time stamp, people knew who conducted certain research first, and thus it reduced the possibility of plagiarism.

In addition to distributing early results, arXiv was used to find new results from other scholars, or researchers with similar interests. An assistant professor in Physics mentioned that "arXiv has been extremely useful and it's been very valuable to get information from the arXiv about who's posting things that are of interest to you" (ID 1839). arXiv also provided long-term accessibility of numerous articles that ended up being published in various journals. One associate professor in Physics (ID 2031) noted that arXiv was a better model to access information than the library where he had to go up and search through the books.

Institutional Repositories (IRs)

The logistic regression analysis in the survey indicated that five factors were positively related to whether or not respondents contributed to the IR in their university (see Table 4.26). Those factors included (1) copyright concerns; (2) accessibility; (3) altruism; (4) trust; (5) rank. Interestingly, I found that respondents who were more concerned about copyright issues were more likely to contribute to the IR. Yet, it was difficult to confirm this finding based on the interview data analysis because most interviewees expressed caution about copyright issues. However, some interviewees noted the issue of trust and accessibility. A couple of interviewees who have not

contributed to the IR questioned the university library's commitment to the preservation of IR materials. This low level of trust in the library made them reluctant to self-archive in the IR. In addition, two other interviewees stated that materials in the IR were not read widely because few in their fields used the IR. This perceived accessibility of IR materials hinders faculty participation in the IR. In this section, I first examine IR contributors' experiences and opinions about the IR. After that, I discuss non-IR contributors' thoughts about the repository.

Seven interviewees identified themselves as IR contributors. They were an associate professor in History, a full professor in Development Sociology, a full professor in Ecology, an assistant professor in Health Informatics, an assistant professor in Sociology, a full professor in Slavic Languages and Literatures, and a full professor in Sociology. Each had different reasons motivating his/her participation in IRs.

The associate professor in History explained that he learned about the IR service at his university, which "promises to maintain documents in perpetuity" (ID 1756). He had conference papers and "non-published, un-reviewed articles" posted on his personal web site, and additionally he deposited those papers and articles into the IR for their long-term accessibility. Nevertheless, he was not entirely sure that deposits in the IR would be preserved like the written materials housed in libraries. Another motivation for his IR contributions was that he wanted to make conference papers public since these had no other outlet for publication. He was proud of the papers and depositing them into the IR was an easy way to make them publicly accessible on the Web. The professor, at the same time, emphasized the importance of a peer review process through which open access scholarly content needed to pass:

> Even though I've circumvented peer review just by publishing my own things, I myself would say they don't really count as knowledge unless some mechanism is put in place to have someone else review it and say that's knowledge. So peer review is still the unsolved problem in online publication (ID 1756).

Similarly, the full professor in Developmental Sociology emphasized the value of long-term accessibility, which the IR offered by noting that "institutional repositories I think are really important because that's a place where there's easy access to your work over a longer period of time" (ID 1803). He expressed two concerns about the long-term accessibility of web resources: 1) the problem of outdated and incorrect professional information on the Web and 2) the difficulty of archiving materials on the Web, which the IR would address. When submitting his work to the IR, he just passed it on with metadata to a library staff member responsible for the deposit. He mentioned that creating metadata was not complicated, although sometimes it was difficult to judge what data was appropriate to enter, for example categories of his work.

In particular, the full professor in Ecology (ID 1507) deposited some of his data sets into the IR. He received a mass e-mail about the opening of the IR from the university library about two or three years ago and he made contributions accordingly. He planned to deposit more data sets into the IR once he finished putting together all the metadata which he submitted as a data paper series to Ecological Society of America. He also planned to self-archive these data sets via a research lab web site for which he worked. He believed that self-archiving data sets in multiple locations functioned as a backup, as well as facilitating access.

The assistant professor in Health Informatics believed that the IR was a place where "you can publish much more than just the written word" (ID 1958), for example, posters, multimedia files, data sets, and software, which were normally not published in a journal. He was also surprised how quickly his materials in the IR were accessible. He mentioned one episode of hiring a computer programmer to work on his research project. He had two final job candidates make a presentation on a pre-selected topic. One of them actually integrated the professor's poster about the research project into his presentation, which impressed everybody. The poster had been deposited in the IR only 6 weeks before, and the professor realized the increased speed of accessing and using materials

publicly accessible on the Web. He suggested that as more professors contributed their research materials to the IR, more use of the information would be made, based on his experience of being invited to an NIH study section. Since he was a tenure-track professor, this invitation was a good opportunity for his career. In the study section, he was told that the NIH people learned about him by locating his poster on the IR. He stated that "I've only been doing this now for about two and a half years and I've already had significant payback from using institutional repositories" (ID 1958).

While the four professors had explicit motivations to participate in the IRs, the assistant professor in Sociology contributed to the IR because her research center was selected as a pilot unit for the university IR. She explained that the Provost's office in her university advertised and supported the IR, and consequently "there was a good deal of outreach and internal press encouraging this" IR initiative (ID 1702). Another reason that she considered her research center suitable to be the pilot unit was the great extent to which people in the research center already self-archived research papers on the Web, and used online resources. As she pointed out, this was what "we do as a collective, that everyone is doing" within the research center, whereas only a few professors in her department made use of the Web to disseminate their research output. In addition, there was a library coordinator who took the responsibility for actually submitting the materials from faculty into the IR. She took care of the whole process of converting research work into PDF files, completing the paperwork, and ingesting it into the IR. This support system made it convenient to contribute to the IR.

One of the IR contributors was responsible for managing a DSpace site for his department. This professor mentioned that he was the only faculty member in the department who had technical knowledge about the IR. He became aware of IRs when he was contacted by the university librarian in charge of the IR initiative. After speaking with the librarian, he became interested in the IR, and decided to start the departmental DSpace site with the librarian's guidance. The materials self-archived in the IR site

included both published work and work in progress created from faculty in the department. Particularly when posting a work in progress, a committee was convened to decide "what should or shouldn't be unpublished there" (ID 1476). The IR site also archived a journal for which the professor was a co-editor. For archiving the journal articles, he and his co-editor recreated PDFs from original files that they used for printing. In the case of other miscellaneous articles, individual authors made PDF files of those. The professor suggested that archiving journals required a series of decision-making processes, as described that:

> In the case of archiving it's a question of which pieces that people want to have archive and electronically available. There are several decisions that go into that – one is whether or not we can easily get copyright permission from the publisher and also the initiative of the individual faculty member to make a PDF scan of the article which takes some labor (ID 1476).

He also commented that his university IR was strongly encouraged by the previous Provost, and without the university support he would not have begun the IR initiative for his department. His departmental colleagues, however, had little interest in the IR so their participation rate was very low. In addition to this inertia, he mentioned that "the whole idea of publishing online is scary" to faculty in his department due to plagiarism or the fear of misuse of their work. Furthermore, the professor noted that he had taught faculty colleagues how to use the Internet, but nobody followed. Older faculty members were even afraid of using the Internet. This fear of technology would be a factor that discouraged the participation in the IR. In order to motivate faculty to contribute to the IR, the professor believed continuing pressure and encouragement generated by universities and funding organizations were necessary.

One IR contributor, a full professor in Sociology who was well-known for a study on children's safety, explained that the only reason he contributed the updated information on the IR was to avoid press interviews. By posting the summary on the IR, the professor could refer reporters to it each time they called or e-mailed him. Like other

contributors' cases, library staff actually deposited the materials into the IR for him. Except for this IR contribution, he has never self-archived his other research on the Web. He even commented that he saw no advantage of self-archiving on the Web, and did not want to learn it because he was old, as described that:

> The world is changing and I don't know. It may be that more and more stuff will get posted down the road but I think that I don't see a great stampede to post things yet, at least in the kind of work that I do. The other thing is I'm getting old. I'm 60 and I don't think that this is a new trick that I'm likely to learn (ID 1695).

While the IR contributors mostly focused on positive aspects of the IRs, three interviewees who have not contributed to the IR cast doubt on the IR services announced by university libraries. Concerning the long-term preservation of IR materials, a full professor in Computer Science mentioned that the university library contacted him to deposit his work into the IR, but since then there had been no follow-up request. He believed that the whole process of the IR initiative was "a fairly low profile activity" (ID 1515). The professor also pointed out that the library needed "a significantly larger investment in the necessary computing and personnel resources than there currently is" in order to preserve materials in the IR effectively. Likewise, another full professor in Electrical and Computer Engineering was uncertain about the feasibility of the IR to fulfill its long-term preservation promises since the library did not have sufficient human and financial resources "to do that for every faculty member who just asks for it" (ID 2235).

In addition to the lack of resources for preservation, one associate professor in Information Retrieval pointed out that IRs were developed without understanding why people would or would not adopt this new technology:

> There are indeed concerns with the long term preservation of the data. The problem that the institutional repositories have is one of managing institutional change. The facilities that they provide for depositing work are not nearly as flexible nor are they as well understood as personal web

> pages. So I think the main problem that they face is managing institutional change rather than the technical issues (ID 1499).

Certainly, a couple of professors stated that the advantage of using their personal web sites over the IR was the ease of controlling their content at any time. They thought that it was inconvenient having IR staff post and withdraw the content. A full professor in Computer Science (ID 1984) did not want to use the IR for self-archiving because his personal web site was already accessed by students and people outside the university.

Professors in Physics and in Mathematics, who have not contributed to the IR, believed that arXiv already covered the need for long-term accessibility of and quick access to their research work. An associate professor in Physics, however, did ask his Ph.D. students to deposit their dissertations into the IR because there was no alternative to preserve such content. In fact, he stated that preserving his work on the Web was "mostly irrelevant" since "in the end nobody's going to care about most of the stuff I have on the Web" (ID 1925).

Two interviewees believed that the IR was not known in their fields, and thus its accessibility was inferior to arXiv. A full professor in Physics (ID 2337) commented that no one in his field posted and read articles in the IR because it was a university web site, which was not an appropriate place for distributing scientific work. An assistant professor in Mathematics stated that "everyone uses arXiv so I don't think [the IR] would increase visibility of the papers at all" (ID 1916). He even mentioned that the only reason that the university wanted to implement the IR was that "the more people that do it the better the university look", and there was no benefit for individuals.

One full professor in Computer Science suggested that the IR would not necessarily be publicly accessible based on his supposition that the IR enabled researchers on campus "to easily access each other's work and also to perform studies like analogous to the one that you are where people who are interested in investigating this creative process will have lots of raw material to work with" (ID 1515). He also

suspected that the university was nervous about copyright holders getting alarmed by making the IR publicly accessible, and consequently the IR activity seemed inconspicuous.

This examination of venues hinted at the motivations for self-archiving through a report on professors' perception of various self-archiving venues. Especially, IR contributors were encouraged by the IRs' promise of long-term preservation, whereas IR non-contributors cast doubt on the IRs' ability to actually do digital preservation, or believed that this function was already fulfilled in other self-archiving venues, such as arXiv. IRs were even considered to be less accessible than other venues since the repositories were almost unknown by peers in interviewees' disciplines. In the next section, I discuss findings from the interviews that specifically concern factors for self-archiving.

Factors affecting Self-archiving

The survey data analysis identified eight factors, which had statistically significant associations with the percentage of self-archived research work (Table 4.29)

Table 4.29. Significant factors identified in the OLS regression analysis

Cost factors	*Benefit factors*	*Contextual factor*	*Individual traits*
Copyright concerns Additional time and effort	Academic reward Altruism	Self-archiving culture	Age Rank Technical skills

These factors are elaborated in greater detail based on the interview data analysis. Since I interviewed 32 survey respondents who have self-archived, and 9 who have not, I compare and contrast these two groups of interviewees. Within the group of self-archivers and non-self-archivers, I attempt to analyze diverse experiences and thoughts about self-archiving practices. In addition, I examine other factors found not to be

significantly related to self-archiving in the survey analysis, as well as some emerging concepts that would be connected to self-archiving behavior.

Cost Factors

Additional Time and Effort

The survey data analysis suggested that if respondents were concerned less about additional time and effort required for self-archiving, the percentage of research work that they self-archived was likely to increase (see Table 4.23). A similar result was shown in the interview data analysis. Eight out of nine interviewees who have never self-archived all worried about additional time and effort, although fourteen interviewees who have self-archived stated that posting materials on the Web required minimal time and effort, described as "seconds", "one minute", or "a few minutes a day".

A couple of other interviewees who have self-archived mentioned that although self-archiving did take time, it was not a great amount. Furthermore, it was worthwhile because of the benefits they received. For example, an assistant professor in Ecology stated that self-archiving was "one of the few ways now that people in academia can advertise what they're doing" (ID 1505). This would make her research recognized better and give her more opportunities to obtain research grants. An assistant professor in Health Informatics (ID 1958), who made IR contributions, also pointed out that the payback from the IR outweighed his time and effort taken for deposit into the IR.

In spite of the potential benefits, self-archiving was not on the priority list for most interviewees, so that updating their personal web sites was often be delayed. One assistant professor in Mathematics stated that "I'm not sitting on pins and needles saying oh by tomorrow that thing needs to be on the web site and I'm not putting too much effort into that" (ID 2457). A full professor in Electrical and Computer Engineering noted that

he tended to postpone maintaining his web site due to his teaching which currently occupied most of his time:

> What's the metaphor- the squeaking wheel gets the grease. The teaching really needs it here and now and posting my research articles on my personal home page is more a thing of that would be good for people but there's no immediate demand or necessity to it so it often doesn't get kept up in the same way (ID 2235).

Although most interviewees personally self-archived, one or two full professors hired someone to maintain their personal or research group web sites. In fact, several interviewees wished to do so if they had resources, in order to minimize time and effort required for the maintenance and update.

Except for one, all nine interviewees who had no self-archiving experience discussed extra time and effort as a major barrier to the participation in self-archiving practices. An associate professor in Mechanical Engineering suggested that "I just don't have time to keep things up-to-date and I find that they're useful to get up there but I tend not to keep them updated" (ID 1919). The issues of time, effort, and technical skills, tended to be addressed together. An associate professor in English (ID 1563) who has been in administration since 1994, stated that she had no technical knowledge about starting up a web page and posting materials there and it would take time and effort to maintain it. Even an assistant professor in Mechanical Engineering mentioned that he was "not very computer literate" (ID 2175) and thus it would take much longer time to maintain his research web site by himself than having his graduate student do it.

Copyright Concerns

In the survey copyright concerns were measured by asking respondents the extent to which they agreed that: (1) they needed to ask permission from publishers to self-archive their work; (2) they might infringe on copyright if self-archiving their work; (3) they could not publish their work if self-archiving it before publication. This factor was found to be negatively associated with the percentage of self-archived research work (see

Table 4.23). Likewise, a majority of interviewees expressed concerns and uncertainty about copyright issues involved in self-archiving.

Most interviewees specified that copyright should be respected and they tried not to violate copyright laws through their self-archiving practices. The primary rationale for honoring copyright was that it protected authors from having their ideas taken by others. One full professor in History mentioned that "copyright is important to protect people's work so it isn't plagiarized, or it's protected in the case of plagiarism I should say" (ID 1467). Another full professor in Electrical and Computer Engineering also stated that copyright "entitles you to put a claim on ownership on intellectual contributions and that's a perfectly legitimate and I don't have a problem with that" (ID 2235).

Yet, assigning copyright to publishers transfers not only the responsibility for protecting authors' works but can also give exclusive economic rights to publishers. A couple of interviewees complained about publishers making excessive profits. One full professor in Linguistics commented that "we all want to see broad access to scientific information around the world and I'm uncomfortable with profit for publishers taking over from that interest" (ID 2214). A full professor in Mathematics accused some publishers in his field of price gouging, and refused to publish in such journals. He even asked the university library to stop subscribing a journal that he normally used because it was too expensive. Although criticizing the high cost of journals, this professor supported "the way journals are set up now" (ID 2227) meaning the positive impact of the refereeing process on academia.

Interviewees normally transferred copyright to publishers, and they dealt with the copyright matter in different ways when taking into account self-archiving of published materials on the Web. One way of responding to the copyright issue was not to make any published materials available on the Web. One assistant professor in Economics never posted published articles on the Web because he assumed the journals owned the

copyright. He also stated that "obviously they would like to sell their journals and they don't want readers to download everybody's articles freely on the Web" (ID 2251).

In addition, five interviewees who had no experience with self-archiving specified that copyright concerns made them reluctant participants in self-archiving practices. One assistant professor in Mechanical Engineering (ID 2156) commented that journals forced authors to transfer all their rights to them, and thus authors were unable to post published articles on the Web. Two non-self-archivers also believed that those who self-archived their publications on the Web did not respect copyright as much as they should, as described by another assistant professor in Mechanical Engineering that:

> The only thing that's preventing me from posting full text of journal papers for example is because of copyright restrictions so I know many people do that but technically I think it's illegal and so I tend to be on the very honest side and I try not to violate any of the copyright agreements (ID 2175).

A couple of other non-self-archivers were uncertain about the extent to which publishers allowed for them to make published materials accessible on the Web. An associate professor in English stated that even if self-archiving became technically feasible for her, she was still "very confused about the degree to which [she is] permitted to do this" unless she could "get [her] questions about copyright answered" (ID 1563). One assistant professor in Sociology also commented that "I don't know what my rights and responsibilities are with respect to my published material and so I don't want to run afoul of the law by having that freely available" (ID 1470).

Only five interviewees actually checked publishers' policies to see whether they were permitted to self-archive published articles or books. One assistant professor in Mechanical Engineering did not post her journal article on the Web when she found the publisher did not allow it. In addition to posting published articles, she pointed out that posting early versions of papers required caution because certain journals considered self-archived drafts already published, and consequently did not accept them as manuscripts.

The professor commented that "the copyright stuff can be a little tricky just because different journals have different policies" (ID 1435). An assistant professor in Physics mentioned different self-archiving policies that he knew, for example, *Nature* permitted authors to post their articles 6 months after the publication, whereas another journal named *Clinical Review Letters* allowed immediate postings on the Web. He followed what each publisher required for self-archiving.

This case-by-case decision making for self-archiving was also carefully made by another assistant professor in Sociology (ID 1702). She usually looked at requirements described in publishers' web sites. On occasion, she asked publishers directly about self-archiving policies. In some cases they answered in the affirmative and in other cases in the negative. When in doubt, she asked the library coordinator who was in charge of the IR submission for clarification. A full professor in Slavic Languages and Literatures (ID 1467) also stated that he checked publishers to get written permission when posting publishers' PDFs.

Besides journal articles, one full professor in Computer Science asked permission from publishers to post some chapters of his textbooks on the personal web site. He mentioned that his publishers were lenient toward self-archiving a few chapters and examples on the Web. Ultimately, however, he wanted other people to purchase them because he "put a lot of effort into it and there's no need to just give it away to anybody that wants it" (ID 1984).

Interestingly, five interviewees were aware of the possibility that they could negotiate with publishers before signing in the copyright agreement form, for the non-exclusive right to copy and distribute their published articles. Three of the interviewees only recently recognized this option and thought about it for self-archiving later on. Two full professors in Developmental Sociology and in Slavic Languages and Literatures noted that there was a reasonable way to "publish without giving up your copyright" (ID 1803) so that it was possible to "reproduce ones' own work without having to beholden

to the publisher to get permission" (ID 1467). One assistant professor in Mathematics was explained this option by his department chair who "amends the copyright forms before he returns them to the publisher and they've never made an issue out of it" (ID 2109). This assistant professor mentioned that "in the future I could null and void that part of the copyright form", which restricted self-archiving, and would make his published articles available via the Web.

Moreover, two interviewees checked whether publishers allowed them to have non-exclusive rights when selecting journals in which to publish their research, or before signing the copyright agreement form. An assistant professor in Health Informatics preferred to publish in journals that allowed authors to retain non-exclusive rights, whereas his collaborators did not consider it as much as he did, and thus he has often been "outvoted". He asserted that current copyright laws were very restrictive and accordingly hindered "the advancement of science and the speed with which research results get from the bench to bedside to the community" (ID 1958). Another assistant professor in Molecular Biology requested non-exclusive rights prior to transferring the copyright to publishers, as described that:

> Typically I'll ask the publisher…when they ask me to sign a copyright form, I'll say yeah I'm fine with all the restrictions you place on it in so far as you allow me to post this on my web page and distribute it to colleagues as I see fit and students. And I've not had one yet say no (ID 2115).

Nevertheless, this negotiation process with publishers was perceived to be daunting by two other assistant professors. One assistant professor in Mechanical Engineering stated that "if they have a big enough name that that journal wants their publications, wants articles from these people to be published in their journals so they're willing to waive those rights" (ID 1976). He dare not ask for the rights from publishers because "a small guy like myself just starting out I can't bend their arms". The other assistant professor made a similar comment that "I don't feel like I'm in too strong of a

position to negotiate" (ID 2457), even though he knew of one prominent mathematician who did so.

Three professors who self-archived published articles did not check whether publishers permitted self-archiving, or just ignored their policies. A full professor in Physics (ID 2337) found it so annoying to ask permission from publishers to post his articles on the Web that he decided not to ask and left it up to the publishers to police his site. Similarly, one assistant professor in Mathematics stated that he and his colleagues were never asked by publishers to delete their postings once they were published. He mentioned that "given that they don't complain I just leave it", but "if the publisher told me I had to take it down then if it's published I would because I adhere to their copyright rules" (ID 1916). He perceived that t individual self-archiving behavior was difficult to track down and thus publishers and other copyright owners would "turn a blind eye" (ID 1515).

As explored earlier in the 'Teaching Materials' section, the decision to self-archive teaching materials depended on whether they contained others' copyrighted content. An assistant professor in Mathematics specified that some of his lecture slides included images and tables from a textbook and he was "not sure how legal it is to put some of that stuff on the Internet when they're not my figures or whatever, it's more from the publisher" (ID 2403). An associate professor in Mechanical Engineering expressed concerns about infringing on the fair use rules in copyright law when self-archiving teaching materials. She stated that "if I posted my teaching materials on the Web I have permission to use materials in my classroom through fair use laws. As soon as I make my teaching materials more broadly available outside of my classroom then I could be risking violating copyright laws" (ID 1919). Moreover, an associate professor in Social Statistics suggested that the fair use rules were "totally inappropriate for the current state of technology" in the sense that:

> It's legal for students to come to the library, take an article that I've put on reserve, and make copies for themselves. It's not legal for me to make a copy and post it on the course web site so that students can download it and print it for themselves. The difference between those two I think is trivial but apparently it's a legal difference (ID 1862).

In addition to the copyright of publications, an assistant professor in Computer Science (ID 2238) noted issues of copyrights and licenses relating to software. He made software that he developed publicly accessible on the Web, and hoped that it was useful for scholars, as well as for companies he valued so that they might incorporate his software into their products. He believed that this would be a great opportunity to facilitate the impact of his research. In this case, however, there was a problem of a license and a copyright with the software. Accordingly, he preferred to use the BSD (Berkeley Software Distribution) license on his publicly available software because software under this license can be integrated into proprietary software products.

The Creative Commons License was recognized by eight interviewees, although only two of them have employed it. An associate professor in Physics (ID 2031) used it for his personal photographs, whereas an assistant professor in Health Informatics put the license on his materials deposited into the IR. He mentioned two reasons for utilizing the license. One pertained to the fact that the license made it clear that he was fine with other people using the IR materials and they did not have to contact him to get permission. The other reason was that using the Creative Commons License indicated "a way to express your support and participation in open access activities" (ID 1958).

In sum, how faculty responded to copyright varied between two extremes – no self-archiving not to violate copyright agreement and self-archiving without checking publishers' policies. Only a small number of interviewees were aware of retaining non-exclusive rights and the Creative Commons License, which allowed faculty to self-archive within a legal boundary. These findings imply that university libraries could provide services that guide faculty members to ensure how to manage copyright in self-archiving practices.

Benefit Factors

Academic Reward

The survey data analysis indicated that if respondents perceived self-archiving research materials on the Web as less harmful, for tenure and promotion, the more likely they would be to self-archive (see Table 4.23). This result is consistent with findings in the interview data. Seven self-archivers mentioned that self-archiving had a positive impact on tenure and promotion. Twenty interviewees, including four non-self-archivers, believed that self-archiving had no effect or relevance on academic promotion whatsoever. Among them, six self-archivers specified that self-archiving was neither counted for or against tenure and promotion. No one, however, mentioned any negative consequences for tenure and promotion from self-archiving. Thus, self-archiving is seem as neutral or positive.

The positive impact of self-archiving would be made through the increased citation rates, the chance of receiving good recommendation letters, and an increase in name recognition. An assistant professor in Physics suggested that self-archiving would make his research more cited and as a result, it contributed to his tenure case. He stated that "it's very important because the more people that cite your work the better so your citation record is extremely important for getting tenure" (ID 1839). He also mentioned that showing all his work on the Web would help him receive good recommendation letters from someone who did not know him. In the same vein, an assistant professor in Ecology commented that self-archiving made her research better recognized by other people writing recommendation letters, as explained that;

> I think that promotion and tenure relies on people writing favorable letter that say that your work is significant so whatever you can do to make your work better known for two other people it should be helpful and in hopefully being successful in getting tenure (ID 1505).

Another assistant professor in Mathematics mentioned that self-archiving would make his tenure and promotion easier, even though "it's not as good as a paper that has been published but it's better than nothing" (ID 1916).

Two other tenured professors also agreed that self-archiving increased research reputation and name recognition, and it eventually had a positive effect on tenure and promotion. An associate professor in Information Retrieval stated that self-archiving helped other people understand his research and develop opinions about it, and thus this would "positively impact reputation" (ID 1499). A full professor in Developmental Sociology pointed out that "having wide name recognition that might come about because of your work being present on the Web would be positive for tenure and promotion" (ID 1803).

Two full professors, one in Computer Science and the other in Physics, noted an indirect influence of self-archiving on tenure and promotion. The Computer Science professor stated "certainly in my department if there's a professor who doesn't have a personal web page for instance that's viewed as highly suspicious" (ID 1515). Thus every professor he knew either untenured or tenured, posted papers on their web sites. The Physics professor mentioned that "in my field there are no young people who don't do that" and self-archiving would help "in order to capture people's attention". Yet, he warned that self-archiving without refereed publications "might backfire indeed when promotion or tenure comes up" (ID 2337).

However, twenty interviewees believed that self-archiving research work had no influence on tenure and promotion because tenure and promotion committees were "primarily concerned with publishing in refereed journals or publishing books with recognized publishers" (ID 1862). One associate professor in Mechanical Engineering who have not self-archived any of her research or teaching materials, stated that self-archiving was completely irrelevant on the academic reward system, as described that:

> I don't think my tenure and promotion committee would have any idea how to acknowledge that contribution. They tend to count publication and count money and probably in the other order and counting web pages is really not relevant (ID 1919).

A full professor in Mathematics also noted that letters of recommendation were the other factor on which tenure and promotion cases were decided. He agreed that self-archiving research work made it easier for senior researchers to see all the work, although he pointed out that self-archiving was not the only way to learn about the research. He said that "often times the letter writers who already know the work anyways because they've heard about it or read some of it or maybe they even collaborated or the collaborator advised the student if it's a student or if it's a former student of theirs" (ID 2227).

Out of the 20 interviewees, six noted that there were no positive or negative consequences for tenure and promotion from posting research work on the Web. An assistant professor in Computer Science mentioned that self-archiving "I'm not expecting a positive impact. I would be surprised if the impact would be negative. The impact might be very close to zero also" (ID 2238). Another assistant professor in Sociology made a similar comment that "with respect to tenure there's no advantage to doing it. There's no disadvantage. It's not looked down upon but at the same time it's just something more you're doing on the side" (ID 1702). One associate professor in History noted that self-archiving used to be viewed as a waste of time for junior faculty. Currently it would not count against tenure and promotion but not count toward it either "unless someone decides to risk their career testing that case" (ID 1756).

The pressure of tenure made assistant professors focus solely on publishing in quality peer-reviewed journals so that they created a good record of publications. An assistant professor in Mathematics commented that "I haven't been submitting to online journals or sort of open-sourced kind of places just because I'm concerned about getting tenure" (ID 1966). They would like to make their research work publicly accessible on

the Web after getting tenure since then they would not be " bound by these rules of getting tenure" (ID 1976).

In addition, one assistant professor in Video Arts explained a tenure system in his field different from that in other disciplines. In order to get tenure, his video art works needed to be exhibited by a museum, and this was the accepted "form of peer review" (ID 2012) in his discipline. He told a story that when he went to the third year review at his previous institution, he made a public video art project on his web site, which had around 10,000 visitors and were viewed more than 800 hours. Although these numbers were considered more than those in most museum shows, the project was not counted in the same way that a museum exhibition was regarded, because his video work was displayed only on the Web. He mentioned that the guideline of tenure and promotion in the institution indicated that "publications in both book form and in electronic" did count, but nevertheless it was "not clear exactly how it [electronic work] will be weighted" (ID 2012).

Long Term Accessibility

In the survey, the long-term accessibility was not examined as a separate factor, although it was found to be the top motivating factor that would affect faculty IR contribution in the future (see Table 4.16). In fact, the long-term preservation of publicly available materials on the Web was a concern expressed by twelve interviewees. Among them, four informants mentioned that the permanence of their web-based content was difficult to ascertain for several reasons: the ephemeral nature of the Web, the obsolescence of file formats, and university technical services no longer available after they would retire. Concerning the volatility of the Web, two interviewees noted that changes in URLs hindered locating certain documents on the Web. One assistant professor in Molecular Biology stated that "URLs were pretty plastic. You can put one up and links break and pretty soon nobody has access to it again" (ID 2115). This was "the

femoral nature of the Web", which made web-based content "less useful, less important, not useful in the very long run" (ID 2090). In addition, one full professor in Mathematics mentioned the problem of file format obsolescence. He speculated that in 20 years all scholarly material might be online but a new format for making them accessible might be developed. Then old formats would not be readable anymore. Relating to this issue, he told a "funny story", as described that:

> If you want to look up records from 100 years with [the city where he lives] it would be very easy – you just go to the office and you look up the records. No problem. Go look up records from 30 years ago. It's very hard because they're kept on a computer in a format that they can't read anymore. So there is a downside with making things accessible online only because the format requirements might change and it's no longer possible (ID 2227).

Another full professor in Electrical and Computer Engineering was skeptical that his course web sites, for example, would be preserved after he retires. He commented that "the institution will say oh he's no longer here and therefore assume that the information is out of date and not feel that within their responsibility to maintain those and keep them there" (ID 2235). He also noted that he could find journal articles published 200 years ago in the library, whereas he was "not 100% confident that the stuff that's posted on the web will be there for another 40 or 50 years".

Journals or publishers were considered to play a role in archiving scholarly works by six interviewees. An assistant professor in Mathematics stated that journals were very committed to preserving published articles and expected to archive them in perpetuity (ID 1916). A full professor in Computer Science suggested that this archiving function that journals provided was the reason that "the Web is not a replacement for archival journals" (ID 1515). One assistant professor noted that he would rely on journals for preserving his research after he retires. Prior to that, his web sites would be maintained by his university.

In fact, two other assistant professors mentioned that they had little concern about long-term accessibility for materials on their personal web sites because the universities would maintain them securely. An assistant professor suggested that "the university will be there forever so basically whatever I put up on my web page will be accessible forever" (ID 2238). Another assistant professor in Ecology also stated that she trusted the function of her university server because "it gets backed up daily so there is hopefully a fair bit of security and longevity to that set up" (ID 1505).

A couple of other assistant professors, however, were frustrated by the limited file storage space that their universities assigned to them. One Economics professor explained that she set up a teaching web site and posted lecture materials there, but at a certain point her "quota went over" (ID 1999) so that she had to delete material from the web site. Another professor in Sociology stated that she had audio/video Power Point presentations that she wanted to self-archive on her web page. Yet she had a concern about "the amount of space that I have on my own space that I have with my own institution for posting on their server" (ID 1702). She thus, ended up giving the presentations to her funding agency, which provided enough server support to post them.

Furthermore, several interviewees suggested that OA repositories including arXiv, IRs and the PubMed Central made a commitment to preserving scholarly content on the Web. Three professors in Physics stated that arXiv played an important part in archiving research in their disciplines. Specifically, a full professor mentioned that arXiv mirror sites existed in Los Alamos and in Germany other than one in Cornell, and those were unrelated servers. This made sure that files in arXiv were maintained in a secure fashion. Four interviewees also regarded IRs as a place that would preserve web-based content in the long run. An assistant professor in Health Informatics stated that depositing research work into the IR was a much safer way of preserving content than keeping it on his laptop. He also mentioned that the PubMed Central would archive his work as well since it had "the resources of the NIH behind them" (ID 1958).

Overall, faculty members had various opinions about who would be responsible for long-term accessibility of digital scholarly content. Some interviewees emphasized the archiving role of scholarly journals, whereas others viewed OA repositories as entities in charge of digital preservation. Faculty members also had self-oriented perspectives on long-term accessibility of their research content, in respect to risks against preserving research works and retention periods.

Ease of Access

The factor of accessibility in the survey was measured in terms of two aspects: the extent to which respondents agreed that (1) self-archived materials were easily accessible through Internet search engines; (2) self-archiving would increase the chance to communicate their research findings to peers. This factor was found to be positively related to the IR contribution, although it was not significantly associated with the percentage of self-archived research work. Yet, the interview data analysis showed that one of the benefits from self-archiving with which numerous interviewees agreed was that it enabled a wider audience to gain quick and easy access to scholarly content. In this section, I examine interviewees' perception of ease of access to self-archived materials. This is followed by a discussion about interviewees' thoughts and experiences with communication to peers through self-archiving practices.

Twenty-seven interviewees including five non-self-archivers remarked on this benefit factor. An assistant professor in Mathematics stated that "it's a wonderful thing to have available online whether it's my work or someone else's – it's so much faster – I love it" (ID 2109). One associate professor in English who has never self-archived her research or teaching materials mentioned that self-archiving "makes it much easier for people to get a hold of" (ID 1563) scholarly work. In particular, a full professor in Developmental Sociology highlighted the value of self-archiving that allowed wide dissemination of his research. He denoted that his ultimate goal of research was to make

an impact on "public decisions about immigration" (ID 1803). In this respect, self-archiving made it possible to reach "a broad public audience that was made up of the general public but also decision-makers, key decision makers especially in my state".

Specifically, the accessibility of research work resulting from self-archiving was believed to facilitate the advancement of science. A full professor in Linguistics illustrated that he used to travel across countries to see particular works when he was a graduate student. He went to a conference there and arranged an extra day to go to a library and use a book that he did not have easy access. This whole process was "very cumbersome, very awkward and time consuming" (ID 2214). Compared to the past, he could now have easy access to historical manuscripts, for example, once those were publicly accessible on the Web. He commented that this was a great advantage to the research process and thus, self-archiving ultimately promoted science. Another full professor in Ecology who self-archived his data sets also mentioned that providing the Web accessibility for his data would eventually contribute to science, as described that:

> The science is really not of value to other people unless it's accessible. So I think one of the best ways to make my data and my interpretations accessible is to post them somewhere…. We've gotten to the point where so many people around the world have access to the Internet that that kind of communication is becoming increasingly important in science in a way of sharing and publicizing ideas and data (ID 1507).

Communication to Peers

Self-archiving was found to be helpful for sharing ideas and exchanging comments among colleagues. Fourteen interviewees believed that self-archiving helped them have contact with other people either whom they knew already or did not know in an efficient manner. Six of them suggested that comments from peers about their research posted on publicly accessible web sites were very helpful. An assistant professor in Health Informatics stated that "all the time I get e-mails and different people see my work and they offer comments and suggestions" (ID 1958) and this was one of the benefits

from self-archiving. Another assistant professor in Mathematics believed that self-archiving increased communication to peers, noting "a couple of things that I put up somebody e-mailed me the next day and said oh that looks really interesting and we talked a little bit about the work" (ID 1916). Yet, he was usually contacted by people that he already knew and talked to about the research. Thus, only posting research on the Web did not always result in comments from colleagues.

In particular, three interviewees mentioned that self-archiving early versions of their research papers generated comments from other researchers, which helped correct the papers prior to submitting them to a journal. A full professor in Linguistics noted that self-archiving was "very common in Linguistics like in a lot of fields to want to get as much input as possible before you submit a paper to a journal" even though this was less than was done in arXiv. Two professors in Physics who have posted pre-prints in arXiv, stated the benefit of getting "immediate feedback from people who have done similar work or they have a comment so you can tell immediately if it's good work or not worth as much" (ID 2452) before sending it to publication. Besides, one Mathematics professor commented that self-archiving his papers enabled him to receive comments, correct mistakes in his work, and post the corrections even after publication:

> One or two papers of mine were wrong. It happens. Even I've got one or two published papers that have mistakes in them so you put a correction up on the Web. That's one good thing about doing it online because you can always post a correction. You can say this particular proof is wrong and maybe post a correction. Or somebody else proofed it the correct way (ID 2227).

In addition, self-archiving made it possible for professors to be contacted by people interested in their research. Eight interviewees noted that they occasionally received e-mails or phone calls from people whom they did not know but found their research work publicly accessible on the Web. A full professor in Computer Science (ID 1515) stated that he was occasionally asked questions by graduate students at other universities who read his papers. Similarly, another full professor in the same field (ID

1984) received inquires about his research on the Web from people around the world with whom he was not acquainted. A full professor in Electrical and Computer Engineering (ID 2090) was also contacted by some people who found data sets on his research group website and became interested in the data. An assistant professor in Physics (ID 1839) mentioned that students looked at his research web site and often came to him asking for jobs. Companies sometimes contacted him to request consulting. Yet, one assistant professor in Economics (ID 1999) commented that she had received e-mails from people in Australia or graduate students in other institutions, although it was difficult to tell how they learned about her research. She assumed that they found out about her by reading her papers on the Web, however, she was not completely sure.

A full professor in Slavic Languages and Literatures stated that contacts from people who he did not know were in fact "an advantage to the online dissemination" (ID 1467), especially communication with people outside of his field. Nevertheless, only a couple of professors reported such experiences. One assistant professor in Mathematics (ID 2457) stated that he received an e-mail from a teacher in California who located some of his teaching materials on the Web and asked questions about it. A full professor in Psychology also noted that "I do know that I've gotten parents of children with disorders or parents who have questions contact me because they've found my stuff publicly posted" (ID 1457). Still, he underscored the increased communication to peers that he knew by self-archiving his teaching materials. He explained that "it was my surprise that when my lecture and course materials were publicly available I found out a lot of my colleagues were checking to see how I was teaching my courses as I did them so it was actually kind of a nice network" (ID 1457).

In contrast, six interviewees suspected that self-archiving itself did not increase communication to peers or people unknown by them. Two interviewees suggested that verbal communication with colleagues was still essential to share ideas. One assistant professor in Mathematics mentioned that his communication to peers has been in "verbal

conversations with people at conferences", rather than "through e-mails". He noted that "that's a fairly standard way to communicate with people" (ID 2109). A full professor in Electronic and Computer Engineering made a similar comment that "the primary ways that I find interaction and contact with other people is not through web publication – it's through conversation, world of mouth, mostly, that sort of thing" (ID 2235). The associate professor in Social Statistics noted that self-archiving did not increase his communication to peers because he always did "trade things by e-mail to targeted people" (ID 1862). The only material he self-archived was chapters of his published book on the publisher's web site. A full professor in Mathematics (ID 2112) also stated that he self-archived unpublished textbooks only and thus it did not affect his communication to peers much.

Use of Self-archived Materials

It is noteworthy that seventeen interviewees stated they have often used self-archived content on the Web for their research or teaching, and they indeed were beneficiaries of content publicly available on the Web. This users' perspective of self-archiving was not explored in the survey data analysis.

Professors who have used self-archived content highlighted the ease of access as a primary reason for the usage. Four interviewees specified that looking through self-archived research papers was the most convenient way to recognize a certain topic in their fields, even though they could find them in alternative ways. A full professor in Slavic Languages and Literature commented that he looked for CVs of other researchers and sometimes the CVs included links to their own papers. He believed that "it's a very quick way to get us into what problems are being worked on in a given field at a given moment" (ID 1467), while pointing out that it was not the only process of understanding what was being studied in the field.

Another full professor in Computer Science noted that if he had some idea, he looked for some of the leaders in the area and checked their web sites, which provided links to their journal articles and other research work. He stated that "it's really important for doing science that people have web pages like my own" (ID 1984). An assistant professor in Mathematics who had no experience of self-archiving also commented that using self-archived content on the Web was helpful to other people who looked into similar problems. She thus employed the Web extensively to generate ideas for teaching and research. She mentioned that "a lot of the research I do is through the databases in our library but I also often look for things that I find in the library databases and see if I can find an online version or a pre-print version by the author" (ID 1966).

Similarly, a full professor in Electrical and Computer Science mentioned that when he prepared a proposal on some topics that he needed to understand more completely, in addition to using databases to locate journal articles, he often searched Google and it helped him clarify the ideas, as explained that:

> I'll just use the Internet search approach usually Google just because it's there and convenient and I'll start looking at some of the web pages that come up in terms of explanatory material. And in enough times it's actually been quite helpful. You have to sort through stuff that's not useful or even inaccurate but nevertheless there's enough good stuff out there and it's been quite helpful (ID 2235).

Google was utilized by several other interviewees to access self-archived content. Two non-self-archivers stated that by putting in keywords or names of particular scholars in Google, they immediately found research papers and downloaded them. They appreciated those who posted their research or teaching materials on the Web, which were very helpful.

In addition, a full professor in Linguistics explained that he used primary source materials publicly available on the Web as a teaching resource for his students. It constituted "new pages of an ancient language in Germany" discovered recently and those who found the pages made them a PDF and it was posted on the Web with a big

news announcement. The professor stated that "20 years ago I would have had to wait years for a journal to bring out an article on this manuscript pages" (ID 2214). Within days, however, he was able to see the manuscript images, marked the most interesting points and send the annotated PDF to his students.

Two professors noted that since they were helped by other people's self-archived works, their own materials on the Web would help others. An assistant professor in Ecology went to other researchers' web sites very frequently and read their papers. She stated that "I find it very useful looking at other people's web sites so I would hope that works in reverse" (ID 1505). One associate professor in Physics also mentioned that "I certainly have been able to use other people's web sites to figure stuff out so hopefully they've been able to use mine to do the same thing" (ID 1803). A full professor in Developmental Sociology appreciated the ease of access to materials of other scholars and this benefit made him interested in posting his own work on publicly accessible web sites. He furthermore mentioned that he was using the Web as a research tool to locate "materials both published and not formally published", as well as data from "publicly available databases" (ID 1803). This reciprocal benefit of easy access to research content existed among scholars. The idea of reciprocity in the altruistic intention to self-archive is discussed in the following section.

Altruism

The survey data analysis indicated that respondents, who agreed more with statements regarding their own altruistic intentions to self-archive, were more likely to self-archive research content on the Web (see Table 4.23). In the survey, altruism was measured by the extent to which respondents agreed with four statements: (1) they would continue self-archiving even if others in their fields did not; (2) they support the principle of open access; (3) self-archiving helped other researchers build on research; (4) self-archiving allowed other scholars to access research works that they could not otherwise.

The interview data analysis suggested that the participation in self-archiving practices was driven more by ideas of reciprocity, rather than pure altruism. Scholars wanted to make their research widely accessible and self-archiving was one way of satisfying their needs but as one professor in Electrical and Computer Engineering specified, the reciprocity among scholars that self-archiving represented was :

> A quid pro quo – if I do it maybe they'll do it and then we can share better. So this whole idea of sharing it kind of fertilizes new ideas that precede it. When people really share their data and their results we tend to make more scientific progress more quickly (ID 2090).

Nine interviewees mentioned that self-archiving allowed people in developing countries, for example, to access research articles that they could not use otherwise. A full professor in Linguistics stated that there was a problem with access to the information (ID 2214). As a journal editor, he sometimes received manuscripts from authors in Central Asia or Eastern Europe who did not have easy access to North American publications. For such people, self-archived scholarly content would be very useful providing that they had Internet access. One assistant professor in Sociology commented that self-archiving was "built in a gap between, built in an equality in who has access to what information… it should be extended into the academy more for that public service" (ID 1470). Similarly, an associate professor in History noted that self-archiving was "more kind of as a public service than as a service to researchers" (ID 1756) and it was for people from outside his field or the country who visited his web site.

Six interviewees mentioned that self-archiving helped other researchers and it would be a main reason for them to self-archive. They hoped it would help other scholars build on their research. This perception, however, was not widely shared among the interviewees. Most interviewees were uncertain whether their self-archived content really benefited other researchers. Yet, one assistant professor in Economics (ID 1999) mentioned an instance in which a Ph.D. student at another university built on the Economic professor's self archived research when writing one of the chapters in his

dissertation. The student asked her whether he could share her data and she allowed him to do so.

In addition, two assistant professors stated that they supported the principle of Open Access – free and unrestricted access to research for all users. An assistant professor in Health Informatics suggested that "my philosophy is I like to have as much open as possible because that's the way to really foster collaboration and idea generation, and advancement of my field" (ID 1958). Another assistant professor in Molecular Biology supported the movement toward OA and he refused to review papers in publishers that he believed "predatory in their pricing" (ID 2115).

Moreover, four interviewees considered self-archiving as an obligation of researchers at universities. An assistant professor in Mathematics stated that "research is supposed to serve the public good and I think that making it available so that other people can read it is part of the public good" (ID 1916). He believed that researchers should have "an obligation to make as much of their work publicly accessible as possible within the confines of copyright laws". In the same vein, an assistant professor in Economics noted that researchers should all feel "an ethical obligation" (ID 1999) to make research publicly accessible to help create and share knowledge. Two other professors discussed this obligation since they were professors at public universities. A full professor in Developmental Sociology stated that "as being a professor at a public university with a public mission it's important that my work be available broadly to the public" (ID 1803). One assistant professor in Mathematics made a similar comment that "I'm at a public university and I'm paid work on problems by the community and so I feel like the community should have access to the ideas that I'm paid to work on" (ID 1966).

Besides, three professors expressed satisfaction on their self-archiving behavior if it helped other people's learning and research. A full professor in Mathematics who self-archived his unpublished textbooks on his web sites stated that "I'm just delighted if anybody can use it. I really am. If I can put something there that somebody can benefit

from that would just make me real happy" (ID 2112). One associate professor in History also commented that "I'm happy to do it… Kind of the Golden Rule dictates that I should do that for the benefit of other scholars" (ID 1756). An assistant professor in Health Informatics mentioned that one of the benefits of being open was "the satisfaction that I get to see that other people are getting benefits out of my work" (ID 1958).

The factor of altruism was observed in several aspects – generalized reciprocity, support for OA, obligations of researchers, and enjoyment in helping other people. In particular, the idea of reciprocity was mentioned by professors who utilized OA content that helped for their own research or teaching. The benefit from using OA materials would be reciprocated to anyone who employs research works self-archived by the faculty members.

Early Dissemination

In the survey, I asked the extent to which respondents agreed that self-archiving allowed for the earlier dissemination of research findings. This item was used with an additional two items to measure the factor of publicity, found not to be significantly related to the percentage of self-archived research work. Yet, nine interviewees discussed their opinions about early dissemination of research through self-archiving.

Earlier versions of research work were made publicly accessible on the Web mainly due to the following two advantages: (1) to receive feedback before publishing and (2) to claim an idea as one's own. The first reason was discussed in the previous section, relating to the fact that feedback from peers is one form of communication among colleagues. There I discussed how three interviewees mentioned that early dissemination allowed them to receive comments about their papers before submitting them to journals.

In regard to establishing a claim for ideas as one's one, three additional interviewees discussed their thoughts. A full professor in Electrical and Computer Engineering suggested that:

> If you're working on something in which it's a horse race and… if you don't get the chance to plant the flag first as soon as you have the idea or come up with a discovery or something… I can see where being the first to put it on a public web site is an advantage rather than waiting for the cycle of time of journal publication (ID 2235).

At the same time, however this professor pointed out that self-archiving earlier work on the Web to claim it first was "pragmatic and perhaps undesirable response to the market place driving forces of academic recognition and credit for ideas that we compete for". An assistant professor in Physics also noted that benefits of early dissemination existed in his field because "you don't want to be scooped so you don't want to wait too long on materials that you've produced" (ID 1839). A full professor in Mathematics explained that the distribution of research via the Internet was so fast that it was easy to know who worked with what topics. He noted that "it's very rare nowadays that two people independently prove something because when somebody proves them they immediately post it and then everybody else can't claim" (ID 2227).

In certain disciplines, however, early dissemination of their research was perceived not to bring forth much benefit since the disciplines were not as quickly changing as hard sciences. An associate professor in History stated that "I know that the fields like physics that it's important to have the data up the next day. History as you might expect moves a lot more slowly" (ID 1756). Similarly, another associate professor in Social Statistics noted that his discipline was "not that kind of fast moving field" (ID 1862) so that he did not see benefits of early dissemination. One assistant professor in Economics mentioned "at least in our areas I feel like none of it is so earth shattering that everybody is striving for the same goal and trying to scoop somebody else. It seems to be more self-defined topics". Another assistant professor in Mathematics and a full professor

in Computer Science suggested that the importance of rapid dissemination was area-specific in their disciplines although in most areas early dissemination did not matter significantly.

In addition, careful review and validation of research work before self-archiving was considered to be more important than distributing it quickly on the Web. An associate professor in Physics agreed the benefit of early dissemination, whereas he pointed out that if the research included novel scientific results, it had to be reviewed internally by a committee in his large collaboration group prior to self-archiving it. Otherwise, "people would look at it with a lot of suspicion because they would know it hadn't gone through the official review process and so at some level it gets less weight than something that has" (ID 2031). One assistant professor in Molecular Biology stated that he did not see much benefit of early dissemination in his field because in his discipline "the review process really is what's going to validate the work so until it's gone through peer review and actually published then the ideas being out there early probably aren't that helpful" (ID 2115).

Trustworthiness of Web-based Content

The survey contained three items regarding trustworthiness of scholarly content publicly accessible on the Web. The items asked the extent to which respondents agreed with the quality of (1) self-archived materials from authors in prestigious institutions; (2) self-archived materials from well-known researchers in their fields; (3) peer-reviewed articles publicly accessible on the Web. The total score of these items was used to measure the perceived trustworthiness and this factor was found not to be associated with the percentage of self-archived research work. However, the interview data analysis suggested that the quality of scholarly content on the Web was an issue that some interviewees worried about and they considered the peer review process essential for quality control.

Five interviewees were wary about the credibility of scholarly content publicly accessible on the Web. An associate professor in Mechanical Engineering with no experience of self-archiving expressed frustration with the difficulty of identifying original authors of research materials on the Web. She believed that it eventually hindered the credibility of the resources, as explained that:

> It troubles me a little bit that you can Google something and find a hundred copies of the same word and have absolutely no idea who wrote those words because they're identical in 100 different pages and very few of those web pages bother to cite who the original author is and it's impossible to tell at this point. And so I find that frustrating because it limits the credibility of that material because it's lost who the original author is and whether it's validated or true or carefully research or from a credible source (ID 1919).

Controlling the quality of web-based content was also perceived to be problematic. A full professor in Food Science who has never self-archived stated that "the fundamental issues that modern information systems are going to have to deal with is garbage in is garbage out and in the modern world is so much easier to get the garbage in – screening is much less well done" (ID 1813). He said that he just finished reading his student papers "pulling material from the Web without having any sense of the quality of the material".

Two associate professors in Physics were also cautious of the quality issue. One of them stated that "there's a lot of stuff out there that's rather dubious in its quality" (ID 1925), although he believed that people could find very good idea if having enough knowledge about what they looked for. The other Physics professor stated that "it's a little scary where you see lots of things out there where the work is not well done" (ID 2031). He suggested self-archivers to be mindful of "where it is that they're putting their work and the fact that they're making it accessible by putting it there". He also pointed out that those who used web-based content always needed to check its source and thought about what they read.

Another full professor in Electrical and Computer Engineering stated that "what's on the Web is good stuff and bad stuff simultaneously and you can hardly tell apart and so it's only viewed as a sort of second class citizen to regular archival journal publications" (ID 2090). He thus believed that peer review was "very fundamental" to maintain the quality of scholarly content. Likewise, four other interviewees noted the significance of a refereeing process. A full professor in Electrical and Computer Engineering mentioned that a peer review process helped researchers realize mistakes or missing points in their work and correct them before making it publicly accessible, as explained that:

> What's the quote I recently heard or read and I liked a lot -- "everyone's entitled to their own opinion, but not everyone's entitled to their own set of facts". And so how do you establish or weed out opinion or what you would like to be the facts versus what really are? And part of that is by verifying that your ideas are self-consistent with those that are more or less proven, the weight of evidence over the year. And I think at least in my field science it's necessary to preserve that filter through which stuff that's made publicly available is vetted (ID 2235).

Similarly, a full professor in Mathematics stated that "the refereeing process is important to keep us honest and obviously… first of all it assures people that papers are correct" (ID 2227). A full professor in Slavic Languages and Literatures believed that "there needs to be some authoritative body that decides what does and doesn't get published and what form" (ID 1467) when making research publicly accessible on the Web. He stated that for web-based scholarly content, "the same vetting processes that print media needs to have" would still be necessary. In addition, one associate professor in History commented that the quality of research materials on the Web was not guaranteed without a peer review process. He however, suggested that it "doesn't mean they shouldn't be out there. It's sort of like Wikipedia. It's great and you learn a lot from it but you know you can't fully trust it. And I would like to think that my conference papers are at a higher level than some Wikipedia but I can't claim that just because I believe it" (ID 1756).

Accordingly, a quality assurance mechanism was fundamental in self-archiving practices, and how to implement such mechanism would be an issue for self-archiving venues.

Citation Rate

Professional recognition was measured by the extent to which respondents agreed that (1) self-archived work would be cited more frequently and (2) self-archiving would increase respondents' visibility in their fields. I grouped those two items as one factor to represent academic recognition and this factor was not significantly related to the percentage of self-archived research work. Still, interviewees did have opinions about the citation rate and visibility and professional recognition of self-archived materials. In this section, I discuss interviewees' perceptions of citation rate affected by self-archiving research content. The next section presents what interviewees believed in terms of the visibility of their research publicly accessible on the Web. I will end this section with interviewees' thoughts about how professional recognition is affected by self-archiving.

While ten interviewees believed that self-archiving their research work would increase its citation rate, thirteen mentioned that they had no idea about citations to their self-archived research. Out of the ten informants, two assistant professors – one in Sociology and one in Economics - specified that their papers publicly accessible on the Internet were cited more than their other papers not on the Web. The Sociology professor stated that her preliminary papers on the Web were "more often cited than the article that's been published" (ID 1702) since the preliminary papers were easily accessible. The Economics professor mentioned one instance in which it took three years for her paper accepted by a journal to be published. Since it took "forever to get published", she decided to make the accepted version publicly accessible on her web site. She found it "cited a few times before it was actually published" (ID 1999) and thus believed that self-archiving would increase the citation rate.

A full professor in Computer Science even commented that "papers don't get cited unless they're available online" (ID 1515). He was more likely to cite papers publicly accessible on the Web because he could easily locate and read them, and understand whether they were relevant to his work. An assistant professor in Economics (ID 2251) mentioned that he heard that self-archiving increased the citation rate from other people who intended to do so, although he did not know it for sure. Interestingly, a full professor in Linguistics stated that he utilized Google Scholar to see citation patterns for research papers on the Web including his own and others':

> I do look at Google Scholar for example and see the certain papers show up as being used by certain people and it's valuable to be able to trace those patterns. So if I'm interested in a particular paper I can go to Google Scholar and see who has refereed to that work and used it so it actually helps me sometimes find additional sources (ID 2214).

Nevertheless, thirteen interviewees noted that they did not know the citation rate of their materials publicly accessible on the Web because they had no meaningful data to calculate it. Two other interviewees also mentioned that the citation rate was not relevant to self-archived research work because they believed that "most of the citations are all for refereed journals" (ID 2031) or "works that's formally published in journals" (ID 1803), not for web postings.

Visibility

Self-archiving was perceived to increase the visibility of one's research and to consequently help them: (1) recruit graduate students; (2) find collaborators; (3) increase the chance to obtain research grant; (4) reach the general public. Four informants stated that one of the reasons for self-archiving was to attract prospective students. A full professor in Physics mentioned that "it's very important because you want to have the best possible students that they get a good impression from your web page to see what research you're doing" (ID 2337). One associate professor without self-archiving experience (ID 1919) stated that although she did not post full-text articles on the Web,

she sometimes post images from her research work or just descriptions of her research in order to recruit new students.

This professor used her web site for developing collaborations and she stated that it was useful for that purpose. One assistant professor in Sociology stated that self-archiving research work did help her find collaborators. Her self-archived materials were located via Google by collaborators whom she worked with, as explained that:

> It's also been very successful in attracting collaborators on research projects so if for example someone is interested in doing a particular project and they're looking for disaster scholars and typing in particular search terms my material will come up and that's been in a couple of cases responsible for people contacting me about my work and leading to some very good collaborations so I haven't had any problem with it (ID 1702).

In addition, it was believed that self-archiving or at least having a research web site allowed researchers to have better exposure in academia and it would increase the chance to have research grants. A full professor in Mathematics (ID 2227) noted that self-archiving made other people aware of one's research work. They might not only be those who wanted to use the work but also people involved in decision-making about awarding federal research grants to him/her. One assistant professor who has not self-archived but had his research web site mentioned that he received funding from companies which found his web site via Google. He believed that "professors that don't have a research web site are at a disadvantage because many people if they want to study a topic they will go to Google and they will put in the name of the topic" (ID 2175).

Moreover, self-archiving was considered to increase the visibility of their work, which was "an outreach to people who are not in [their] disciplines" (ID 1756). A full professor in Developmental Sociology mentioned that since he worked for a public university, it was important that his work "be available broadly to the public" (ID 1803). Another full professor in Psychology also noted that self-archiving "in some way does affect the recognition or access of the general public to [his work] which I think is valuable to the discipline" (ID 1457).

Professional Recognition

The perceived impact of self-archiving on professional recognition differed among interviewees. Eight interviewees agreed that self-archiving would increase academic recognition of research works, although five interviewees mentioned that they did not know or it was difficult to know whether it improved the professional recognition. Four other interviewees stated that they did not expect academic recognition or impact to enhance by self-archiving their research materials.

The eight professors who noted the positive relationship between self-archiving and professional recognition suggested that Web accessibility made self-archived research work available to be read by more people and accordingly this would increase the recognition of the work and the author. Two of them, however, commented that the recognition was only "slightly improved to just having it in the journals" (ID 1435), "not in any dramatic way" (ID 1467).

Three interviewees mentioned that they were uncertain about whether the recognition would be improved through self-archiving or not. Two interviewees stated that it was difficult to measure academic recognition or impact, even though it was measured quantitatively by the number of citations (ID 2238). An assistant professor in Physics noted that "It's hard to measure the impact. I don't have a form that says will you cite our work and has the web site helped you to find articles that are of interest. I mean we don't have any feedback mechanism to check what that impact is" (ID 1839).

In fact, four additional interviewees pointed out that improving professional recognition was not an expected benefit of self-archiving. A full professor in Developmental Sociology mentioned "at the web site we're not really trying to have an academic impact as much as we're trying to have a broader public impact" (ID 1803). Another full professor in Ecology who self-archived his data sets stated that "I don't expect a lot of formal recognition for doing this. That's certainly not a reason why I'm

doing it" (ID 1507). One full professor in Electrical and Computer Engineering commented that academic recognition "has to do with just peer-reviewed archival journal", and thus he found "no advantage – not a lot of recognition" (ID 2090) from self-archiving his research on the Web. One associate professor in History also made a comment that self-archiving would rarely improve academic recognition:

> I would describe it as icing on the cake. I really like it when someone from Jordan e-mails me because he found my article on the Web. It doesn't contribute in a more narrowly professional way to my impact and recognition but it's nice just to make that contact with somebody. So it helps me as a scholar, as an intellectual, as a person but it doesn't help my impact or recognition as a professional (ID 1756).

Besides, a full professor in Physics (ID 2452) mentioned that professional recognition was increased due to the quality of the research work, not just making it publicly accessible on the Web. An assistant professor in Economics also noted that people tended to look for research work of people who already gained academic recognition. She made a comment that "if you were at Harvard or somewhere and had put articles up early that it definitely would increase the visibility of the study" (ID 1999).

Personal Benefits from Self-archiving

Six interviewees suggested that they posted research and/or teaching materials on the Web due to the following personal benefits: (1) easy access to their own "archival material" (ID 1925; ID 2031); (2) reduced burden of responding to e-mails (ID 1505; ID1702; ID 2457); and (3) a feeling of accomplishment (ID 2109). As mentioned before in the 'arXiv' section, an assistant professor in Mathematics described that posting earlier versions on arXiv gave him "some sense of accomplishment" (ID 2109). He stated that this feeling of accomplishment was "personal" benefit from early dissemination of his work, which other people referred to and read without waiting a year for the journal to publish it. These personal benefits were not captured in the survey data analysis.

Self-archiving was also considered useful because they were able to access their research or teaching materials from anywhere. An associate professor in Physics mentioned that since he made all his teaching materials publicly accessible on the Web, he could "quickly look up what it was [he] did in a particular course" (ID 1925). He also stated that he posted grade curves on the Web and those were helpful when he wrote a recommendation letter for a student while being off-site and not having his laptop. Another Physics associate professor regarded his research group web site as "an archive of work that we've done" (ID 2031). He noted that it was "extremely helpful when I'm trying to put together a quick talk that summarizes stuff that's been done… perhaps when I'm traveling". An assistant professor in Sociology commented that it was easier for her to locate a certain material on her web site than in her laptop, as described that:

> I do use it sometimes because it's easier for me to find a file. So I may have a file on my computer because of course I originally posted it to the Web so I must have it but it's sometimes easier for me to download what I uploaded online – that includes the syllabus, that includes the handouts, that includes my own article sometimes to send it on rather than trying to find what folder I've stored it in (ID 1702).

In addition, one assistant professor in Ecology mentioned that self-archiving "helps me to answer people's questions. It makes my life easier. Just go to the web page I can say… I don't have to say it 25 times when I have to talk to all these different people" (ID 1505). Another assistant professor in Mathematics stated that he posted homework problems and solutions on the Web and "that way I get bothered less often by students who miss the class" (ID 2457).

In sum, a variety of intrinsic and extrinsic benefit factors were discussed among interviewees. Interviewees' thoughts about the extrinsic benefit factors were varied, especially about the increased citation rates, which previous studies emphasized. The majority of interviewees stated that they had no idea about citation rates of their self-archived research and thus, they could not discuss it unless having some meaningful data regarding the citation rates. Yet, the interview data analysis supported the positive or

neutral relationship between self-archiving and tenure/promotion. The factor of altruism was also found to be associated with generalized reciprocity, as opposed to pure altruism.

Contextual Factors

Four contextual factors were examined based on the interview data analysis: (1) trust in users, regarding concerns about plagiarism and stealing ideas; (2) self-archiving culture; (3) peer pressure; (4) influence of grant funders and university/department on the decision to self-archive.

Trust in Users

In the survey, trust was measured by the degree to which respondents agreed that: (1) readers might plagiarize or fail to cite their work if it was self-archived: (2) the integrity of self-archived work might be compromised; (3) self-archived materials were not maintained securely. The first and second statements were used to measure the level of trust in users, whereas the third statement represented trust in institutions or other entities that maintained publicly accessible venues. The factor of trust both in users and in institutions turned out to be positively associated with IR contribution. However, it was not related to the percentage of self-archived research work. Based upon the interview data analysis, fourteen interviewees were concerned about other readers using their self-archived work improperly, whereas twenty-three had minimal worries.

In particular, eight non-self-archivers out of these fourteen interviewees who were afraid that their ideas would be stolen if they self-archived research work. Three of them specified that their fields of study were so competitive that they needed to protect themselves from plagiarism. One associate professor in Mechanical Engineering mentioned that obtaining grants and making publications were “very competitive” so that she could not “be as free with the materials that I’m spending my time and my students’ time generating as I might like to be” (ID 1919). An assistant professor in Mechanical

Engineering similarly noted that "a lot of things I read, the things I see people doing… it seems like they've basically stolen ideas and different things and so I'm really worried about that because this business of academia can be very cut-throat at times" (ID 1976). Another assistant professor in Mathematics also stated that "my biggest concern is that other people will appropriate [her research work] and publishing companies will take it and try to copyright it for themselves" (ID 1966). She worried that the companies might use it for profit.

Two other non-self-archivers (ID 2403; ID 2175) had the exact opposite opinion and believed that posting published materials would prevent the problem of stealing ideas. In fact, this was mentioned by six self-archivers who had no concern about the issue. They thought that once published works were self-archived, people could not scoop their ideas because the online publication served as documentation of the idea. An assistant professor in Physics noted that "if the work is published before I put it up on the Web then any plagiarism is obvious because anything that they produce will be done at a later date" (ID 1839).

In addition, it was believed that plagiarism happened even in print media and thus there was no need to worry about it in the Web environment anymore than in print. For example, a full professor in Slavic Languages and Literatures commented that "the danger is always there and my work has been both plagiarized and misinterpreted even in print so I don't think it will make any difference" (ID 1467). An additional two interviewees stated that they understood the risk of plagiarism existed when self-archiving their research work; however, it would be "more valuable to get the information out than it is to be worried about what others will do about it" (ID 2031) because of "the positive benefit of being open" (ID 1958).

Moreover, plagiarism rarely occurred in certain areas of research for several reasons. An associate professor in Physics mentioned that since he worked with very large collaborations, which constituted "highly specialized experiments", it was not

possible that "somebody could just steal the work" (ID 1925). An assistant professor in Mechanical Engineering explained that scooping ideas was not common in her field to worry about because "there's just so many ways to do things and it's not like we're all competing to do the same thing" (ID 1435). A full professor in Mathematics also stated that his field was small enough that "referees would recognize if somebody else tried to plagiarize" (ID 2227). Another assistant professor in Mathematics made a comment that self-archiving on the Web was culturally accepted in his field, and thus there was a hidden rule of giving credit for self-archived research works:

> In Math people in general are not that concerned about other people stealing their ideas. Sort of the quorum seems to be is that once somebody has something written in a decent form then that person has priority. So I say in a way the system works in Math is a little bit of an incentive to getting your work put up in a public place quickly because that way somebody may have the same idea they would have to recognize that you did it first because you have it on the Web and they wouldn't have it on the Web (ID 1916).

Self-archiving Culture

Self-archiving culture was measured by the degree to which respondents perceived self-archiving as common in their fields or their departments. The survey results indicated a positive relationship between self-archiving culture and that the extent to which respondents self-archived (see Table 4.23). The interview data analysis showed that sixteen interviewees including one non-self-archiver perceived self-archiving to be very common in their disciplines, which encompassed Physics, Mathematics, Ecology, Molecular Biology, Computer Science, Information Retrieval, Economics, Linguistics, and Video Art. The single non-self-archiver was an assistant professor in Mathematics. She mentioned that her web site needed to be "reworked because [her] materials have undergone some changes" and "the biggest barrier there is having the time to maintain [her] web site" (ID 1966). She was also "concerned about getting tenure" so that reluctant to submit her research to Open Access venues.

Seven informants including two non-self-archivers mentioned that self-archiving was reasonably common in such disciplines as Sociology, Psychology, Mechanical Engineering, and Electrical and Computer Engineering. However, six interviewees including three self-archivers believed that self-archiving was not common in their disciplines – History, Slavic Languages, Cell Biology, Mechanical Engineering, Food Science, and Health Informatics. The three self-archivers were all IR contributors (ID 1467; ID 1756; ID 1958) and highly motivated to make their research publicly accessible on the Web. One of them also suggested that the participation rate in PubMed Central was very low across fields of Bioscience because investigators were "not aware of scholarly communication issues" and had "lack of knowledge of the benefits to them personally in their own careers" (ID 1958) from self-archiving.

Of the interviewees that discussed self-archiving culture in their fields, three pointed out that there was a discrepancy in self-archiving culture between their fields and their larger departments. An associate professor in Information Retrieval in a department of Library and Information Science mentioned that self-archiving was "so common that there's hardly a paper that I can't get off the Web" (ID 1499) in his field, although it was "uncommon" among colleagues in his home department of Library and Information Science. One assistant professor in Sociology also commented that in her research center, self-archiving was "extremely common" and "almost required" (ID 1702) whereas it was "less frequent" in the department at large. Similarly, a full professor of Linguistics in a German Studies department mentioned that self-archiving was very "uncommon in literary circles – in German literature circles" (ID 2214) but it was "norm now in Linguistics and that does encourage everyone to make the material available" on the Web.

Peer Pressure

In the survey, the influence of three external actors including peers, grant-awarding bodies, and universities or departments was measured by the extent to which respondents thought each actor affected their decision (or not) to self-archive. The survey findings showed no significant relationship of external actors on the decision to self-archive. In calculating this measure, I summed up the scores from all three actors, and thus the effect of each on self-archiving decisions was not observed. Interview data, however, demonstrates individual influence by each actor, particularly collaborators or peers.

Seven interviewees specified that there was a good deal of peer pressure involved in their decision-making processes about self-archiving. All of the seven were from disciplines where self-archiving was perceived as the norm. An assistant professor in Mathematics mentioned that "if nobody else did it I might not do it either" (ID 1916). Another assistant professor in Physics noted that "everybody in my department has their stuff available on the Web so you get some peer pressure" (ID 1839). One assistant professor in Economics also commented that "there's been some movement not to post things when they're under review and I've gotten that from other people. That's how I learn these things so I guess we do what our colleagues do" (ID 2251).

An assistant professor in Health Informatics explained that he decided whether or not to self-archive depending on the majority of his collaborators' or co-authors' opinions. If his collaborators wanted to publish in particular venues, which "preclude us from publishing with an open access venue" (ID 1958), he followed their wishes. In the case where he worked with people who were "like-minded" in a sense that they wanted to publish in OA venues, he would do so. Two physics professors involved in very large collaborations noted that interactions among collaborators to decide what to make publicly accessible were complicated. One stated that a committee within the collaboration developed "collaboration policies that govern accessibility – at what point can results be released and at what point can they be made public, and stages as to how

they're made public and so that is all taken pretty seriously" (ID 1925). The other professor also mentioned that self-archiving was tricky in large collaborations because the data was a "shared resource":

> Many, many people came together to try and produce that data so if you are going to take the data and then publish it and not check with your colleagues first… that would be bad. And the reason is most cases your colleagues will have very good points about your analysis and if you ignore them you do it at your own risk and perhaps are doing less high quality science than you might otherwise be doing (ID 2031).

Another assistant professor in Economics commented that she still collaborated with her former Ph.D. advisor who has self-archived every paper on his web site and this role modeling motivated her to self-archive. Thus without hesitation, she posted papers co-authored with him or made links on her website to those from his web site. However, she felt "awkward" (ID 1999) asking other co-authors whether they would self-archive papers particularly when she was not the first author. She felt that she "had done a little bit less on" the papers and did not want to "somehow make them feel like [she was] claiming the full paper".

Influence of Other External Actors

Other than the influence of collaborators or peers, the interview data analysis showed the effect of grant-awarding bodies on self-archiving decisions. Six interviewees acknowledged some influence by grant-awarding bodies on their decision to self-archive or not. Two professors – one in Ecology (ID 1507) and one in Molecular Biology (ID 2115) noted that NSF (National Science Foundation) encouraged them to deposit their data sets into publicly accessible databases on the Web. One assistant professor in Mechanical Engineering (ID 1435) referred to NIH (National Institutes of Health), which requires grant recipients to self-archive journal articles into PubMed Central. For this reason she made her publications available there. An assistant professor in Sociology (ID 1702) mentioned that her grant awarding body – the Public Entity Risk Institute –

required her research group to send research outputs, such as final reports, case studies, or preliminary papers to the Institute and this regulation was in the grant contract. It then posted the materials on their web site. Another full professor in Developmental Sociology (ID 1803) stated that his funding agency - US Department of Agriculture - emphasized an education and outreach component to their sponsored research. He believed that self-archiving would help to satisfy this funding agency's goal in this area so he posted results from the project on the Web. A full professor in Electrical and Computer Engineering (ID 2090) mentioned that he maintained a large research group web site funded by NASA, which had great influence on his decision to self-archive data sets on the web site.

Three interviewees specified that since they received research grants from "tax payer money" (ID 1984), they had an obligation to make the results of their research publicly accessible on the Web. An associate professor in Physics, however, commented that this reason was not a main motivator for his own self-archiving, he explained:

> I wouldn't say that I make it publicly accessible because of that…I mean I do fundamental research in science and that's the name of the game. So I don't find myself ever thinking about having to get a paper there because of the source of my funding because I'm already trying to get the paper out there. I don't need that additional little push (ID 1925).

Two non-self-archivers suggested that industry sponsors often did not allow them to post research findings on publicly available web sites (ID 1919; ID 2175). An associate professor in Mechanical Engineering also mentioned that if his company-sponsored research included proprietary components. Therefore, he was not permitted to self-archive these results. He described that "as long as it is not proprietary I see no problem because most of it is developed with public assistant, government funding in other words" (ID 2156).

Interviewees perceived little influence from their universities or departments, either encouraging or discouraging self-archiving. Only two professors at the same university, one of whom was an IR contributor, mentioned that they were "encouraged by

the university" (ID 1435) to make contributions to the IR because "the previous Provost made public statements that this was a goal of the university to do more electronic dissemination" (ID 1467). Even with this encouragement, however, only one actually did self-archive. Two assistant professors (ID 2251; ID 2457) stated that other than offering web space and maintaining web sites, their universities played no role in self-archiving practices. In addition, a full professor in Mathematics noted that his department head felt "a little negative about people putting materials on the Web" (ID 2112). The professor speculated that it was helpful for the department head to say that faculty members in his department produced a large number of refereed articles in journals because then the administration of the university would think that there was a lot of research going on in the department and consequently, more research money would be allotted to the department.

The survey data analysis showed that the existence of self-archiving culture was positively related to self-archiving. Along with this, the interview data analysis suggested that there was peer pressure in disciplines where self-archiving was common. In addition, almost all interviewees who had no experience of self-archiving expressed concerns about plagiarism. The fear of plagiarism was even more severe among faculty from highly competitive disciplines. These factors relating to disciplinary norms influenced faculty's decision to self-archive.

Individual Traits

Age and Technical Skills

In the findings from the survey, age and self-rated technical skills were positively related to the extent of self-archiving (see Table 4.23). Eleven interviewees believed that younger professors were more likely to self-archive than older professors. In particular, three of them suggested that younger people were more computer-literate than older

people, and they noticed some correlation between age and technical skills. Technical skills and disciplines also seemed related to one another. Professors in Humanities tended to have less technical skills than ones in Computer Science and other Engineering fields.

However, the distinction between the younger and older generations was made based upon interviewees' subjective judgments, depending on their own ages. An assistant professor in Mathematics (age:39) stated that "everyone that I know who is under 45 puts all their stuff on the web on arXiv" (ID 1916). While this professor specified the maximum age of the younger generation, other interviewees did not. A full professor in Linguistics (age: 52) noted that self-archiving was common among "younger people and people who are very active" (ID 2214) on their research. An assistant professor in Economics (age: 35) commented that it was difficult to answer whether or not self-archiving was common in her field because "there's more of a generational divide" (ID 1999).

Two non-self-archivers suggested that age might be an issue for self-archiving. An associate professor in English (age: 57) suggested that "the whole idea of putting stuff out on the web that's a younger generation than mine" (ID 1563). One assistant professor in Sociology (age: 37) mentioned that "I think some of it is generational that you'll find younger scholars less concerned about putting their stuff up in publicly available ways" (ID 1470). This Sociology professor was even 20 years younger than the English professor and in that sense, what they referred to the younger generation would differ greatly.

Three other interviewees perceived age to be related to technical skills. An assistant professor in Mechanical Engineering (age: 33) stated that "since I'm a starting out professor right I'm probably more computer savvy than some of the guys that have been around for 20-30 years" (ID 1976). A full professor in Mathematics (age: 49) noted that "the older generation some of them feel uncomfortable with the use of computers…but I think people in my generation or younger they generally just put their

stuff on the web" (ID 2227). However, another assistant professor in Mechanical Engineering (age: 34), who was 15 years younger than the Mathematics professor mentioned that he was not very computer-literate, and thus hired a graduate student for maintaining his web site. He said that "I think this new generation of students…they are much better with the computer" (ID 2175) so that they knew several "tricks" that guided posting research on the Web in an efficient way. These interviewees commonly assumed the correlation between age and technical skills, whereas what they perceived to be the younger generation was different.

One professor in Slavic Languages and Literature stated that he was the only one faculty member who had technical knowledge about web posting. The rest of colleagues, especially older professors in his department had some "fear of technology" (ID 1467). In addition, an associate professor in English was reluctant to self-archive partly because she did not "really know how to do it" (ID 1563). However, technical skills were not discussed much by interviewees in other disciplines. Specifically, it was common that Engineering professors built a web site and posted research there by themselves. Two assistant professors in Mechanical Engineering stated that they created their web sites and the preparation of self-archiving was "not that much more" (ID 1976) for them. One assistant professor in Computer Science also noted that "I'm technically savvy enough to do it myself" (ID 2238).

Summary of Interview Data Analysis

The interview data analysis suggests that faculty participants value the peer review process in scholarship. The emphasis on the journal refereeing process leads to an almost universal preference to self-archive peer-reviewed articles. Still, interviewees were interested in self-archiving other types of research materials, but they noted that they often had no channel for dissemination. This finding implies that IR staff need to

embrace a broad range of research material in different formats to meet faculty members' needs, and to develop strategies for collecting and preserving various scholarly content.

Personal web pages were the favorite venue for self-archiving because professors liked the ability to fully control their content. They also noted the fact that research on personal web pages was easily discoverable via Google. Moreover, they believed that university or department servers were secure enough to maintain their work. Accordingly, professors using personal web pages were not particularly convinced by the promise of long-term preservation in IRs. Those who had already contributed to the IR, however, were more enthusiastic and more likely to view the IR as a preservation system. This finding indicates that IR staff need to address digital preservation issues.

Reciprocity was one of the important themes emerging from the interview data. Faculty self-archivers strongly supported the principle of Open Access, and I measured this in the construct of altruism in the survey. Relating to this concept, interviewees stated that they wanted to reciprocate after receiving benefits from using OA content by giving back to other researchers through self-archiving their research. Professors with public funding or those employed by public universities also felt more of an obligation to self-archive in an OA venue whether or not this was mandated. Given the complexity of the affective, legal, technical, and administrative factors discussed, the relationship between reciprocity and self-archiving behavior needs to be further examined in the future.

CHAPTER 5

DISCUSSION

This chapter discusses major findings of this study and how these answer my four research questions. It also addresses the implications of the findings. The first and second questions were (1) what are existing ways that faculty members make research materials publicly accessible on the Internet and (2) why do they use certain forums for self-archiving? Those questions relate to types of scholarly work that faculty self-archive, as well as the venues they use and reasons for that usage. The third and forth questions concern motivating and impeding factors affecting self-archiving, respectively.

Faculty self-archivers prefer to make refereed articles publicly accessible on the Web because their quality has already been assured by peers. Faculty consider peer review essential to control the quality of scholarly work. I suggest that there is a continuum of peer review from career review based on "affiliation or on prior established credential" of authors (Ginsparg, 2002) to the journal refereeing process. This would be a useful concept for dealing with the lack of quality control mechanisms in most self-archiving practices. For example, arXiv employs a career review function - an automated filter on authors' affiliation to accept papers by submitters from recognized research institutions.

In addition to peer-reviewed articles, faculty participants in this study wanted to self-archive a variety of other research materials associated with their papers, such as data sets, graphs, audio and video files, or posters. This potentially valuable scholarly content needs to be properly managed in a sophisticated cyberinfrastruture. If IR staff want to

play in this arena, they will need to amass expertise on metadata creation, develop an infrastructure that is sustainable, create preservation regimes for diverse files over time, and maintain long-term accessibility.

Faculty self-archivers using personal web pages tend to be satisfied with the fact that they have control over their research and at the same time, this research is easily located via Internet search engines. It is not a high priority among faculty to ensure long-term accessibility of their scholarly content. Somewhat surprisingly then, digital preservation was considered to be the most desirable feature of IRs. This is interesting since IRs were not created as vehicles for preservation and many in charge of IRs are not guaranteeing long term preservation. This implies that there is a gap between perception and the reality of digital preservation in the context of IRs. There is also no consensus concerning who in the IR staff is responsible for preservation – library and archival staff are equally likely to have a role in preservation.

Along with the digital preservation, survey respondents who contributed to IRs were significantly more concerned about copyright issues than non-contributors. This implies that IR staff need to provide guidance on copyright management to alleviate the confusion over copyright. In fact, most interviewees expressed uncertainty about copyright regardless of whether they have made contributions to IRs. A few IR contributors, however, recognized the possibility of retaining non-exclusive rights. This indicates that they have better understanding about copyright issues regarding self-archiving.

This study identified several motivating factors for faculty self-archiving. Unlike previous studies that asserted increased citation rates and the resulting impact of OA content as a motivator of self-archiving, this study found that faculty self-archivers were encouraged by the philosophy of Open Access that stresses the benefits for end users. The principle of OA implies altruism, although rather than pure altruism, the idea of reciprocity motivates faculty to self-archive. Previous research often suggested that the

tenure and promotion system discouraged faculty participation in self-archiving. Yet, this study found that faculty self-archivers considered self-archiving to have either positive or neutral impact on tenure and promotion. To what extent self-archiving was common in disciplines also determined faculty self-archiving behavior. Self-rated technical skills indicated faculty's comfort level of technologies, positively associated with self-archiving.

Impeding factors found in this study confirmed previous studies. Major barriers to adopting self-archiving practices include concerns about copyright issues, as well as the time and effort required for self-archiving. IRs may be able to address these concerns by providing advice and assistance of copyright management and mediated deposit services. In addition, concerns about plagiarism were prevalent among those interviewees who had never self-archived. It was found that competitive secrecy tended to be dominant in their disciplines.

Self-Archiving Practices

RQ1: What are existing ways that faculty members make research materials publicly accessible on the Internet?

Importance of Peer Review

Refereed articles were the primary type of research material that faculty self-archived. The value of a peer review process that ensures the quality of research papers persists in the self-archiving environment. As Harnad (2003) suggests, research literature without peer review would be "neither reliable nor navigable, its quality uncontrolled, unfiltered, un-sign-posted, unknown, and unaccountable" (p.338). He emphasizes self-archiving peer-reviewed papers in Eprint archives, which make the vetted research openly available on the Internet.

Harnad also mentions that although journal peer-review systems are far from perfect due to their slowness, subjectivity, and bias, no other feasible mechanism for quality control of scholarly works is currently available. Interviewees in this study who stressed the value of peer review recognized those problems; however, they noted that the certification of research was fundamental for scholarship and they asserted that it would never be changed. King et al. (2006) even suggested that peer review was so deeply embedded in academic value systems and well-supported by journals that scholars would rarely adopted an experimental approach, such as IRs, to disseminate and archive final publications,.

Nevertheless, I found a couple of instances of peer review other than the journal refereeing process in the interviews. One was an internal review process in large collaboration projects mentioned by two professors in High Energy Physics. Before self-archiving research works produced by the collaboration were self-archived, a committee selected from the collaborators carefully reviewed the research. It was easy to find reviewers among participants in the huge scale of collaboration. The internal peer review was as rigorous as the journal peer review, and thus it sufficiently upheld the quality and accuracy of the science. This finding demonstrates how researchers in collaborative projects can create their own quality control mechanism, enabling them to less rely on the journal refereeing system. In addition, members of the collaboration I interviewed were inclined to self-archive a variety of research on the Internet to maximize the accessibility of their research without much concern about its trustworthiness.

The second example of an alternative peer review mechanism was described by the professor in Slavic Languages and Literature who was responsible for managing his department's site in the university IR. He mentioned that the DSpace site collected both published and unpublished materials created by members of his department. Prior to depositing unpublished materials or works in-progress, a committee assembled and decided which work they would self-archive. This instance sheds light on the possibility

of implementing peer review in IRs at the level of a department or discipline (or in the terms of DSpace, a community). At the same time, it raises issues regarding who would be responsible for managing and acquiring this work and how the peer review process would be set up to evaluate its quality, as the Slavic Languages professor has done. He stated that the university IR was strongly supported by the previous Provost. This university-wide endorsement motivated him to take responsibility for managing IR content, along with his interest in the IR. This implies that universities need to make their position in support of the IR services explicit and visible so that faculty members become aware of IRs as a sustainable, reliable, and viable peer-review model of disseminating their research.

These alternative methods of quality assurance indicate "a spectrum of peer review, from peer review 'light' to collaboratively reviewed materials to traditional top-down peer review" (Harley, 2007). Given the digital environment where a variety of scholarly content can be utilized to meet users' needs, the classic peer review system may not always be effective for assessing the quality of individual research material. Technologies also make it possible to automatically screen inadequate research materials posted on the Web. For example, arXiv provides a filter on institutional affiliation of authors submitting papers. The automated filter restricts the arXiv author base to physicists from recognized research institutions or faculty members in major research universities. This career review function provides a minimum criterion to identify submissions qualified to be reviewed (Ginsparg, 2007).

Unlike arXiv, IRs can include many types of research other than peer-reviewed articles and thus IRs have more difficulties in ensuring the quality of their content than disciplinary repositories. Figure 4.7 indicates that the majority of faculty IR contributors deposited pre-refereed articles or unrefereed articles into the IRs. The majority of IR collections examined in the MIRACLE project encompassed traditional text-based documents – theses and dissertations, journal articles, working papers, and pre-prints

(Markey et al., 2007). IR staff need to explore other types of scholarly content valuable to users because research data and output span a number of formats. Two interviewees in this study explained their experiences of depositing content other than textual materials into IRs. Their experience and thoughts give an insight of managing multiple IR content in the future.

Self-archiving Non-Textual Scholarly Content in IRs

Data sets were found to be self-archived by the fewest number (n=107) of faculty participants (Figure 4.1). Compared to paper-type research works, it is not very common to make data sets publicly accessible on the Internet. However, interviewees from Ecology or Molecular Biology were familiar with open data sharing and have deposited data sets into publicly accessible databases. The only interviewee who has contributed data sets to the university IR was a full professor in Ecology. He viewed the IR as a place for preserving his data sets in perpetuity and the promise of long-term preservation convinced him to make the IR contribution. He considered it important to create a "data paper" that described the context of collecting, measuring, and reporting the data set. This metadata description of data was crucial for ensuring its long-term accessibility, because ecological data are highly diverse, and they are used and reused in a number of ways. The contextual information needed to be carefully recorded for current and future user communities (Karasti et al., 2006).

In order to collect and preserve data sets in IRs, it is essential to provide proper metadata description for data. Unlike creating metadata of paper-type research works, data sets can be described the best by data originators who know the most about the data. Yet, the metadata description requires a good amount of time and effort, which discourages data creators to deposit data sets and metadata. Karasti et al. (2006) pointed out that in actual practice of LTER (Long-Term Ecological Research) network, the

information managers often play an important role in addressing metadata issues. They educate data originators the importance of metadata creation for long-term accessibility of their data sets. Information managers also provide data management services and incentives that help data contributors describe metadata. If nothing above worked well, information managers consult various sources regarding data, usually identified from papers written based on the data, and assign the necessary metadata.

Assigning metadata requires a profound understanding about the context of research domain where the data were created, collected, and analyzed. IR staff who manage data sets and metadata need knowledge about a wide range of data sharing practices, to properly contextualize the data. They also need to actively work with faculty members who deposit data sets to provide a rich context of data creation, use and reuse. Zimmerman (2006) suggested that without data repositories with open access, standardized metadata, and quality control, ecologists were able to collect and validate data sets for secondary use by applying formal and informal knowledge acquired from disciplinary training and prior data gathering experience. This approach, however, had a limitation, especially when they retrieved data sets meeting their specific requirements for research. Collaboration with faculty members who create data, IR staff, and IT specialists would be necessary to make data sets accessible in the long-term.

In addition to data sets, supplementary materials associated with a research article – figures, tables, posters, or presentation slides – are potentially valuable research materials for IRs. An assistant professor in Health Informatics was excited about the opportunity that the IR provided to make his posters and presentation slides, as well as graphs and tables that he thought useful but were eliminated through a refereeing process due to page length limitations. He noted a few instances where such materials were actually used by other people shortly after deposited in the IR. He was invited to an NIH meeting, which was an honor and aided his career, because the meeting coordinators located his poster on the Web. These examples demonstrate the IR's capability of making

non-textual scholarly content accessible in an efficient manner. IR staff need to explore this opportunity to collect and manage the supplementary content types, in addition to related research articles.

Moreover, IRs face a challenging issue of managing research articles, underlying data, and other supplementary materials associated with the papers. They are all conceptually related with one another and users may want to see all in one place. In fact, operational IRs with more than 200 documents collected raw data resulting from doctoral dissertations as one of the major content types (Markey et al., 2007). Still, most data and their ensuing articles do not in one archive. The Source-to-Output Repositories (StORe) project explores ways of linking data sets in one repository to research papers in another repository (Polydoratou, 2007). According to Lynch (2007), this new paradigm of connecting scholarly articles to data sets and related digital objects requires sophisticated cyberinfrastructure that enables users to view the associated content all together, as well as changes in policies and behavior that adopt the approach of managing scholarly content. IRs are part of this cyberinfrastucture and need to proactively respond to the emerging way of relating papers, data sets, and ancillary digital objects together.

RQ 2: Why do they use a certain forums for self-archiving?

This study identified the main venues that faculty members used for self-archiving (Table 4.8). Personal web pages were the most popular venue, followed by research group web sites. Departmental web sites were the third most utilized venue, although faculty members who planned to self-archive selected this venue the second most frequently for future self-archiving (Table 4.15). In contrast to this survey finding, only one interviewee self-archived technical reports on the departmental web site. This venue was often used for posting course materials or non-scholarly materials, such as newsletters or pictures of events. A couple of other interviewees mentioned that their personal web pages resided on a departmental server when asked about their use of

departmental web sites. This conceptual overlap between personal web pages and departmental web sites may have inflated the findings on the current and future use of departmental web sites for self-archiving.

The survey results also indicated that IRs were the least favorite venue for current and future self-archiving. Yet, I found 109 actual IR contributors among the survey respondents and performed a logistic regression to identify significant factors affecting their contribution to IRs. Prior to discussing the reasons that faculty self-archive in IRs, I will examine the relative advantages of self-archiving in personal web pages mentioned by the interviewees. These advantages provide interesting lessens for IRs and address what faculty want for the repositories.

Personal Web Pages: Control and Flexibility

The main reason that faculty self-archive in personal web pages is the ability to control research output. By using personal web pages, faculty members are able to update and delete their works at any time. They can also organize self-archived content based on categories that made sense for themselves. They often create hyperlinks to their papers in other publicly accessible web sites, which is simpler than posting them on the web pages. Interviewees noted that they were unwilling or reluctant to give up these advantages by adopting other self-archiving venues, especially IRs, which they perceived as a means of losing control over their content.

Personal web pages also function as a current awareness tool that provides the most recent research conducted by faculty. Google plays an important role in making the research quickly accessible to the public. This venue, however, lacks stable and long-term accessibility of scholarly content. URLs are changeable and breakable so that previously identified documents often cannot be located later. Goodman et al. (2006) noted that out of 11 social science papers self-archived in faculty's personal web sites,

only five were found in the same locations 15 months later. The other 6 articles were moved to other web sites so that previous URLs did not work. The authors were able to locate the articles only by retrieving them via search engines. It was suggested that personal web pages did not sufficiently fulfill the long-term accessibility of research works.

However, interviewees discussing the long-term preservation of their research tended to have a self-oriented perspective on the issue. They believed that their research materials would be useful until they retire, and after that nobody would care. University or department servers were robust enough to store and backup their research as long as they stayed at the university. After faculty members retire, journals would preserve their publications in perpetuity and thus the archiving function of journals was considered to be essential.

Digital preservation of scholarly content was generally not an issue for faculty members, whereas those interested in contributing to IRs in the future were attracted to the idea of the IRs' ability to preserve their materials (Table 4.16). This finding implies that digital preservation should be a core function of IRs, in spite of the fact that there is currently little consensus on the part of IR staff on the extent to which IRs should be responsible for preservation (Hockx-Yu, 2006).

IRs: Long-term Accessibility and Copyright Management

The logistic regression analysis identified significant factors that influenced faculty contribution to IRs (Table 4.26). It implies that IR staff need to develop best practices for ensuring long-term accessibility of IR content, as well as for addressing copyright issues in respect to self-archiving. In this section, I discuss findings relevant to long-term accessibility issues and three practical aspects of digital preservation on which IRs should focus: (1) file formats; (2) preservation metadata; (3) organizational support.

Following that, I suggest a potential role for the IR staff in coordinating resources on campus relating to copyright issues to inform current and potential IR contributors of how to manage copyright of their research materials.

Long-Term Accessibility

Two factors concerning long-term accessibility were found to be significantly related to IR contribution: (1) accessibility of self-archived content and (2) trust in users, as well as institutions responsible for maintaining OA content. IR contributors believed much more strongly than non-contributors, that self-archived content would be easily accessible and it increased the chance to communicate research findings to peers. This belief has been supported by IRs providing permanent URIs for deposits, which addressed the problem of plastic URLs.

Compared to non-IR contributors, IR contributors believed more strongly that users acknowledged their self-archived works properly, and those materials would be maintained securely by institutions. In particular, the perceived trust in preservation of self-archived materials coincides with the finding that faculty members were motivated the most by the promise of long-term preservation in IRs. The results imply that IR staff need to investigate appropriate models and practices of digital preservation for the repositories, and eventually implement IRs as a preservation system of various scholarly contents.

According to Hedstrom (2003), "long-term digital archiving requires systems, institutions, and business models that are robust enough to withstand technological failures, shifting computing platforms and media, changes in institutional missions, and interruptions in management and funding" (p.vii). The social, technological, and economic aspects of digital preservation have not yet been fully investigated. Since little has been known about best practices of digital preservation, preservation is not highly prioritized in the IR development process (Rieh, 2007). Nevertheless, this study

demonstrates that digital preservation is a selling point for IRs to increase faculty contribution. IR staff need to make adopt digital preservation standards and best practices as one of their core functions. Based on the results in this study, I make following suggestions for long-term preservation of IR contents.

File Formats

This study showed that faculty submitted textual content, such as refereed or pre-refereed articles most frequently to the IRs, whereas a couple of interviewees mentioned that they deposited other file formats into IRs, such as data sets, figures or tables. This indicates that IRs must have the ability to preserve not only textual materials, but also images, multi-media, or data files. It is unavoidable that a variety of file formats exist in IRs and accordingly, there would be the wide range of risks associated with file format obsolescence, however IRs that focus on .pdfs will miss out on significant amounts of scholarly output and potential contributions. One interviewee even specified that the threats of obsolescence would be a downside of archiving research in digital form.

IRs should address this concern by assessing preservation risks of each file format based on several factors, which include its wide adoption, history of backward compatibility, good metadata support, a good range of functionality, available interchange format with usable target, built-in error checking, and reasonable upgrade cycle (Digital Preservation Management Tutorial, 2003). Recognizing the risks helps IRs develop policies on recommended file formats, which would be safe options for reducing the level of preservation risks. This is also a good rationale that encourages faculty and students to adopt such file formats when they deposit research works into IRs. However, IR staff need to apply the file format policies with flexibility so that these should not be a barrier to IR contribution. IRs can accept valuable research content in other file formats with greater risks. In this case, it is necessary to make plans for the migration of the formats to ensure long-term accessibility of the content.

Preservation Metadata

The accessibility of IR content was found to be a factor that motivated faculty to contribute to the IRs. OAI-PMH is a mechanism that enables IR content accessible via Internet search engines and thus, metadata play an essential part in enhancing the IR accessibility. A full professor in Ecology who deposited data sets in the IR specified that it was important to provide metadata "where somebody could find the data if they wanted to follow up on your study" (ID 1507). The descriptive metadata facilitate the immediate search of IR content, although long-term accessibility is supported by preservation metadata, which provides information needed to make digital content readable, viewable, and interpretable over the long-term.

DSpace IRs employ Dublin Core (DC), and its elements mostly represent descriptive information. Administrative and technical information that support long-term preservation are not present in the DC standard. The information necessary for preservation is described in Preservation Description Information (PDI) of the OAIS reference model: (1) reference information, which describes identifiers of digital content; (2) provenance information, which explains the history of the content; (3) context information, which illustrates the relationship between content and its environment; (4) fixity information, which documents authentication mechanism that prevents content from being altered (Digital Preservation Management Tutorial, 2003). A standard such as PREMIS comes closest to articulating a full set of preservation metadata elements.

IR staff need to identify a common set of preservation metadata and begin to collect them along with descriptive metadata. Metadata librarians can collaborate with university archivists and system administrators, and utilize their expertise to develop the process of gathering preservation metadata. Archivists who have knowledge about electronic records management help identify provenance and context information necessary for long-term preservation, as well as technical issues. System administrators

might assist IRs to collect technical metadata regarding file formats, software and operating systems.

Organizational Support

A couple of interviewees expressed doubts about an IR's ability to preserve research content in perpetuity because they believed that university libraries did not have sufficient human and financial resources for long-term preservation activities. This perception contrasts considerably from my finding concerning IR contributors' trust in library commitment to digital preservation. In fact, digital preservation programs reside in an organizational context and IR staff need to assess resources and needs in the university community to fulfill long-term preservation of IR content.

Elements of organizational infrastructure in Trusted Digital Repositories (TDR) can be used to evaluate responsibilities in both a higher organizational level and an operational level. Those elements include (1) governance and organizational visibility; (2) organizational structure and staffing; (3) procedural accountability and policy framework; (4) financial sustainability; (5) contracts, licenses, and liabilities (CRL-OCLC, 2007).

According to the criteria, university libraries need to provide a mission statement that reveals their commitment to long-term preservation of IR content. A formal succession plan or contingency plans might be necessary in case the IR ceases active operation. IR staff demonstrates appropriate skills and experience regarding legal, technical and archiving issues to accomplish their responsibilities. In addition, it is important to develop written policies and procedures and keep them updated when technology and community practices change. Making comprehensive documentation of all relevant preservation activities would be helpful for establishing systematic approaches to IR digital preservation. Ongoing funding is also essential for digital preservation and libraries need to move away from project-based funding, such as grant

or gift. Long-term and short-term business plans need to be developed and reviewed periodically. Moreover, IRs provide deposit agreements for contributors to notify that they agree with transferring certain rights necessary for preservation activities in IRs. IRs should make continuous effort to satisfy requirements of TDR to build trust in the repositories, which in turn, motivates faculty to contribute their research to IRs.

Copyright Management

Interestingly, the logistic regression indicated that professors who are more concerned about copyright issues are also more inclined to contribute to IRs. This finding implies that faculty IR contributors consider university libraries as making a commitment to manage copyright for IR content. One interviewee who has contributed to the IR noted that she was helped by the university librarian responsible for the IR to identify the self-archiving policies of journals.

It is necessary to create IR policies that specify rights and responsibilities of authors, as well as services that give guidance to faculty on copyright issues. A reasonable goal of this activity is to help faculty members and library staff manage copyright by articulating copyright issues in evolving scholarly communication systems, rather than to enforce copyright compliance or to provide legal advice to staff or users. IRs thus, play a role in coordinating human and information resources relevant to copyright on campus to raise awareness of the issues and to develop guidelines of copyright management.

Raise awareness of copyright issues

Within a university library, IRs can share the responsibility for copyright management with other units, for example, scholarly publishing office, digital library service program or the university legal counsel. Staff members in such units could share knowledge and expertise on copyright issues. To assess the state of knowledge and

existing practices, IR staff need to perform an internal environmental scan by conducting a survey and having meetings with those with expertise in copyright around the university. This activity could lead to greater collaboration between librarians others with knowledge about copyright.

Outside the university library, IRs could build partnerships with legal counsel and university policy officers leading to more consistent policies concerning intellectual property around the university. IRs formulate policy statements and check legal languages of the policies. In addition, providing educational opportunities for faculty and students would increase their awareness of copyright and intellectual property issues. There may be individuals or groups in a university who have expertise in copyright and they can be invited to a series of colloquia or panel discussion regarding copyright and Open Access. By organizing these educational forums, library staff can increase the visibility of IR services to university members and they would recognize that IRs are willing to address copyright concerns.

Develop guidelines of copyright management

Among interviewees, concerns about and confusion over copyright issues were prevalent no matter whether they have contributed to IRs or not. To minimize the uncertainty of copyright management in self-archiving practices, guidelines or IR policy statements need to be provided. The type of collaborations I have suggested could form the basis of more transparency in reviewing existing copyright and intellectual property policies and statements used on campus or in other institutions.

In this study, only a few interviewees realized the possibility of retaining non-exclusive rights and this indicates that copyright retention was not widely known by faculty. The guidelines should inform faculty of how to retain a certain rights needed for self-archiving. This study also showed that awareness of the Creative Commons license was very low. Only 13.5% (n=92) of survey respondents were aware of this type of

license. Of those who were aware of it, a mere eighteen faculty members had used the license. I suggest that librarians be more assertive about educating faculty about Creative Commons licensing and other models that enable faculty retain copyrights so that they know available options and choose what they prefer.

Motivating Factors

RQ 3: What motivates faculty members' self-archiving behavior?

Altruism: Support for the Spirit of Open Access

Interview data suggest that faculty members were more encouraged by generalized reciprocity in self-archiving than by true altruism. This "generalized reciprocity" occurs when knowledge is considered as a public good in online communities. This form of sharing knowledge represents a "generalized exchange", one of the social exchange structures (Molm, 1997). In generalized exchange, resources are transferred to one actor, but the one who reciprocates is not the actor directly receiving those resources. Seventeen interviewees stated that they have used self-archived research work, which they were able to quickly access via Internet search engines. They benefited from OA content and appreciated those who shared it publicly. A few of the interviewees even commented that they would like to return this favor by self-archiving their own research.

I contend that faculty self-archivers accept the idea of sharing their research as a public good and this perception leads them to develop altruistic intentions to disseminate research publicly on the Internet. In the present study, the factor of altruism was found to be positively associated with the percentage of self-archived research materials. Three interviewees, however, described their motivations for self-archiving as an indication of pure altruism. Moreover, another four interviewees were inclined to self-archive because

it was a moral obligation, they also cited their employment as professors at public universities as an important consideration. These findings conform to results of previous studies, although they are not consistent with what Harnad (2006) argues – from authors' perspectives, self-archiving is a matter of self-interest in increasing impact of their research. Still, in my study faculty self-archivers are concerned with making their research as widely accessible as possible, and in that sense their altruistic intention is not entirely free from self-interest.

Positive or Neutral Impact on Tenure and Promotion

Survey respondents who perceive self-archiving as having a less harmful impact on tenure and promotion, tend to self-archive more research work. All the interviewees who mentioned the tenure and promotion process remarked that there was either a positive or neutral relationship between self-archiving and academic reward. This finding contradicts other studies indicating that the current tenure and promotion system impedes faculty member's adoption of newly introduced models of publishing and disseminating research (King et al., 2006; UCOSC, 2007).

Previous studies conflate OA publishing and self-archiving into new models of scholarly publishing. These studies then conclude that faculty do not want to change their behavior because these new models lack peer review and prestige, which the current tenure and promotion system emphasizes. In my study, these concepts were not conflated. Asked about self-archiving separately from the OA publishing paradigm, I found that faculty self-archivers understood self-archiving as a complementary addition to distribute their research. They continued to research and write for conventional publications, valued by the academic reward system, yet viewed self-archiving as having little effect on tenure and promotion.

In fact, seven interviewees noted positive effects of self-archiving on academic rewards. Two assistant professors mentioned that self-archiving raised the recognition of their research and they thought this helped generate good recommendation letters during their tenure reviews. The remaining interviewees also stated that self-archiving increased the likelihood that other researchers would read and discuss their research, and as a result, this would positively influence name recognition and reputation. Although this study found that the perceived increase in recognition or impact of self-archived research was not a motivating factor, some faculty members speculated that there was an indirect effect of the factor on tenure and promotion.

Self-archiving Culture and Peer Pressure

In this study, self-archiving culture was operationalized based on faculty members' perceptions of whether or not self-archiving was common in their disciplines. I found that this perceived culture of self-archiving differed among the four areas studied – Science, Engineering, the Social Sciences, and the Humanities. The average score of self-archiving culture was highest in Engineering, lowest in the Humanities, and in between in Science and in Social Science (Table 4.31). Yet, this study also revealed differences in self-archiving culture within different fields in these areas (See Appendix E). For example, overall Engineering professors generally perceived a strong self-archiving culture; however, interviewees in Mechanical Engineering mentioned that self-archiving was moderately common or even uncommon in that field. Interviewees from Computer Science agreed with the existence of a self-archiving culture. In the Science area, interviewees from Physics and Mathematics perceived self-archiving to be widely accepted, although professors in several sub-disciplines of Biology (Molecular, Ecological, or Cell Biology) had diverging perceptions regarding self-archiving culture. The result demonstrates disciplinary variations in self-archiving culture.

Although professors in the Humanities view self-archiving as atypical behavior in their fields, only 10.6% (n=51) of total self-archivers identified in my survey belonged to Humanities (Table 4.9). Interestingly, out of the 125 respondents from the Humanities, 40.3% were self-archivers. While 40.3% may seem large, this proportion was almost half that reported by those in other areas (Table 4.11).

Some interviewees from disciplines with a strong self-archiving culture, such as Physics or Mathematics, reported peer pressure to self-archive their research. In addition, new self-archivers learned conventions of self-archiving, e.g., not posting papers under review, by looking at what other self-archivers in their disciplines did. The influence of peers or collaborators may reinforce the adoption of self-archiving as well as specific practices for self-archiving in such disciplines.

Technical Skills, Age, and Rank

The three individual traits - age, rank, and technical skills – were found to be significantly related to the percentage of self-archived research work. While age was negatively associated, rank and technical skills were positively related to the intent to self-archive. In other words, younger professors, tenured professors, or professors with more technical skills tended to self-archive a greater percentage of their research.

The association between age and self-archiving implies that younger professors are more familiar with using the Internet for disseminating their research than older professors. In fact, several interviewees assumed that younger faculty would have more technical skills so therefore would be more inclined to self-archive. However, the distinction between younger and older generation was based on interviewees' subjective judgment, usually in reference to their own age. Thus, the exact age of those 'younger' professors who are likely to self-archive is relative.

Technical skills are also related to the disciplines to which faculty members belong. The average score of self-reported technical skills is greatest among Engineering professors, followed by faculty in Science, Social Science, and the Humanities (Table 4.30). This finding indicates that professors whose domain knowledge involves computer technology, such as Computer Science, are much more likely to participate in self-archiving practices.

In addition, professors with higher rank tend to self-archive a greater percentage of research work. This result implies that tenured professors are more inclined to self-archive than untenured professors. I suggest two rationales: (1) tenured professors may have more research materials for self-archiving than tenure-track professors and (2) tenured professors are relieved from tenure concerns so that they may be more willing to try new publishing models than untenured professors. The latter reason aligns with the UC faculty survey which suggests that senior faculty would be potential adopters of new initiatives in scholarly communication, as opposed to junior faculty (UCOSC, 2007).

Impeding Factors

RQ 4: What makes them reluctant to self-archive their research materials?

Confusion over Copyright Issues

Three interviewees who have not self-archived believed that they had no right to post their research articles on the Web because they assigned exclusive rights to publishers. From their perspective, most self-archivers violated copyright law and the non-self-archivers did not want to be involved in illegal activity. Two additional non-self-archivers were unsure of journal policies regarding self-archiving and thus they decided not to self-archive.

Uncertainty about copyright and the responsibility of authors was commonly mentioned by self-archivers. Yet, only five self-archivers regularly checked self-archiving policies on journal web sites or through e-mail contacts. They understood that since the journal policies varied, they needed to be cautious when deciding what to make publicly accessible on the Web. Five other self-archivers noticed that it was possible to request an amendment of copyright agreement in order to retain non-exclusive right for self-archiving. Among the five self-archivers, however, only one actually requested and modified his copyright agreement. Two assistant professors stated that they felt uncomfortable negotiating with publishers because they worried that their request would endanger future publishing opportunities with that publisher.

Overall, faculty members tried to respect copyright and believed that their decisions concerning whether to self-archive or not was made within a legal boundary. However, non-self-archivers were more concerned about copyright restrictions than self-archivers, who understand that a certain level of flexibility exists in managing copyright for self-archiving.

Concerns about Additional Time and Effort

While most interviewees who actually self-archived noted that self-archiving required minimal time and effort, a few self-archivers stated that it did take time but the benefit from self-archiving prevailed over the amount of time and effort that they put in. A couple of other self-archivers also mentioned that since self-archiving was not highly prioritized, maintaining and updating their personal web sites was often delayed.

Thus, the issue of time and effort is not only important for non-self-archivers, but for most faculty members. Nonetheless, professors who have not self-archived do not see the advantages of self-archiving and believe that the advantages outweigh their time and effort. The level of technical skills that they possess is also related to the effort to learn

the logistics of self-archiving. This study lends support to Davis and Connolly's (2007) suggestion that non-self-archivers avoid the learning curve for self-archiving practice unless they perceive it adds value.

Fear of Plagiarism

Eight out of nine interviewees who had no experience of self-archiving were concerned about the possibility of getting scooped if their research was self-archived. In the survey data analysis, the factor of trust, which partly measured the concern about plagiarism, was found not to be significantly related to self-archiving. Nevertheless, all but one non-self-archiver in interviews explicitly worried about plagiarism, and thus I added the fear of plagiarism as a disincentive for self-archiving. In particular, three non-self-archivers mentioned that their fields were highly competitive so that they did not make research publicly accessible until it was published. According to Walsh and Hong (2003), scientific competition strongly predicted secrecy, which the authors measured in regard to "unwillingness to discuss ongoing research with those outside the research group". Scientific competition was operationalized based on "concern over having one's research results anticipated" (p 802). This is related to the non-self-archivers' reluctance to self-archive their research because given the competitive circumstance, self-archiving research in-progress would increase the risk of getting scooped. Consequently, the non-self-archivers kept their ongoing research private up to the time of publication.

Based on the discussion of the motivating and impeding factors, Figure 5.1 illustrates the relationships of the significant factors to self-archiving.

As seen in this figure, factors found in interviews are depicted as boxes with dotted lines, while the remaining factors were identified in the analysis of the survey. The relationship of the factors to self-archiving is indicated by a positive or motivating (+) or negative or impeding (–) sign. Some factors are related but their directional association is

unclear, for example, the relationship between disciplines and technical skills. In such cases, factors are connected with lines.

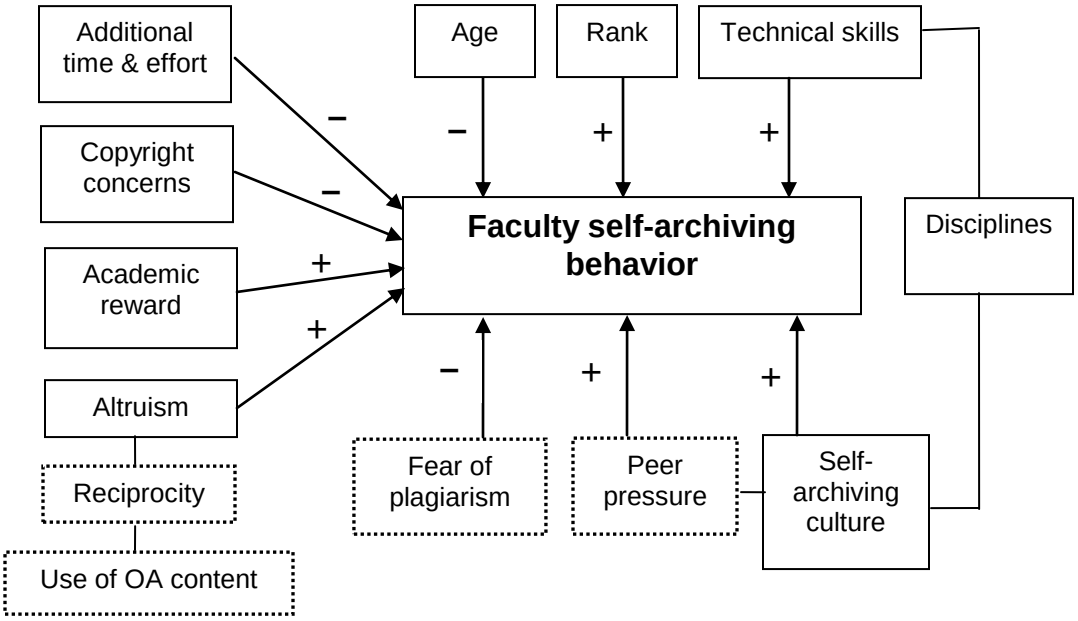

Figure 5.1. Updated Model of Factors affecting Faculty Self-archiving

Compared to the original model that I proposed in Figure 1.1, the updated model includes several additional factors: (1) reciprocity; (2) use of OA content; (3) fear of plagiarism; (4) peer pressure; (5) age; (6) technical skills. Except for the fear of plagiarism, these factors were rarely examined in previous studies. Unlike the original model, the new model also specifies directional relationships between each factor and faculty participation in self-archiving practices. The updated model demonstrates that various social factors and individual traits affect faculty's decision to self-archive.

CHAPTER 6

CONCLUSIONS

This chapter explores the implications of the results of this study from both the theoretical and practical perspectives. This study is based on two theoretical discourses – (1) the Socio-Technical Interaction Network (STIN) model that Kling et al. (2003) proposed and (2) Social Exchange Theory (Hall, 2003; Kankanhalli et al., 2005). I used the STIN model to build a general framework for this study and then inserted Social Exchange Theory into this model to specify motivating and impeding factors on self-archiving. This study broadens the scope of research driven by the STIN model and Social Exchange Theory, in identifying social and technical aspects relevant to faculty self-archiving practices. My findings also indicate that digital preservation and copyright management are essential IR services for encouraging faculty participation in IRs. In conclusion, I will discuss limitations and future research directions identified as a result of this study.

Theoretical Implications

STIN Model

According to Meyer (2006), the STIN model has not reached the level of theory, nor is it a proper methodology. He asserts, though, that it is an "analytic strategy" that

facilitates finding "more complete explanations and thorough understandings of the relationship between the social and the technical in socio-technical systems" (p.44).

This analytic approach helped identify two aspects of socio-technical networks involved in self-archiving: (1) existing communication forums; and (2) resource flows. In order to communicate research, faculty members primarily use traditional scholarly publishing systems. These existing communication regimes, such as peer-reviewed journals, are deeply embedded in the academic values and reward system. Accordingly, professors rarely opt for innovations for scholarly publishing, which they perceive do not support the functions of traditional publications – quality control, distribution, and archiving. While self-archiving meets faculty members' needs to disseminate research as widely as possible, professors have cast doubt on the quality and the long-term accessibility of self-archived research. In that sense, faculty members believe that self-archiving supplements the current scholarly publishing mechanism as a new means of distributing research.

Self-archiving alone does not lead or increase communication to peers. Faculty members utilize the existing communication channels to inform peers of their research in advance and self-archive research papers later on. They are then contacted by peers, receive comments, and exchange ideas. Some interviewees were also contacted by strangers, but often cannot tell how these individuals learned about interviewees' research, because the research is distributed via both conventional publications and self-archiving. Therefore, the effect of self-archiving on the communication of research is difficult to explore because of the multiple scholarly communication regimes simultaneously involved in the dissemination process.

Kling et al. (2003) suggested that diverse resources including money, skill, and status are one of elements in socio-technical networks. Following the resource flow, especially money, helped identify any impact on interactions within these networks. The present study examined the influence of grant-awarding bodies and the university or

department which provides resources and possibly exerts control on faculty members, upon their decisions to self-archive. According to the interview data analysis, the university or department had little influence on faculty self-archiving decisions, whereas six interviewees mentioned that their grant-awarding bodies encouraged them to self-archive research output funded by these agencies. Particularly those who received money from public institutions felt obliged to self-archive so that the general public could access the research findings. In contrast, two interviewees funded by companies noted that self-archiving was prohibited if the research contained proprietary components. In these respects, there is some evidence that the attitudes of grant funders toward self-archiving influence professors' decisions to (or not to) self-archive.

Social Exchange Theory

Incentives and disincentives for self-archiving are particularly related to Social Exchange Theory, which suggests that the costs and benefits resulting from the exchange of resources influence the decision to share resources. In this study, resources are operationalized as various types of research material created by faculty members. Social exchange theory has been applied to information sharing activities in corporate environments, but it has rarely been used for analysis of information exchange in scholarly communication settings. This study demonstrates its applicability to self-archiving activities, and the ability of social exchange theory to predict significant factors, and thus conceptually contribute to research on Open Access.

Self-archiving represents a generalized exchange process where benefits received by an actor are not usually contingent on resources provided by that actor. I found that professors who have used scholarly OA content appreciated the ease of access, and in return wanted to self-archive their research. Some of these faculty members even felt obligated to make their research publicly accessible on the Internet because their research

was funded by taxpayer money, or they worked in public universities. Unlike earlier research , however, my research indicates that self-archiving behavior tends to be ruled by reciprocity, rather than by pure altruism. According to Cropanzano and Mitchell (2005), reciprocity might be considered as a universal principle whereas it is unlikely that all individuals value reciprocity to the same degree. In general, people with higher endorsement of reciprocity were more likely to return good deed. It would be interesting to examine the relationship between self-archiving behavior and the extent to which faculty members apply reciprocity principles.

Practical Implications

Among various self-archiving venues, IRs are implemented in university settings and libraries play a key role in IRs. Findings from this study indicate that long-term digital archiving and copyright management are the two major areas on which IR staff need to focus. Although university libraries are not totally unfamiliar with these areas, increasing contributions by faculty may depend on libraries taking more initiatives and determine sustainable and extensible strategies for implementing services in these areas.

Digital preservation policies and functions in IRs have not yet been well-established. The majority of operational IRs surveyed by MIRACLE project implemented policies regarding acceptable file formats, authorized contributors, access restriction, metadata formats, and acceptable content (Markey et al., 2007). However, preservation policies need to embrace not only technical issues, but also the organizational context where preservation strategies are developed and implemented. The high-level organizational responsibilities of Trusted Digital Repositories mostly lack in current IR policies on preservation. Organizations should understand (1) their local requirements, (2) other organizations who might share responsibilities with them; (3) which responsibilities can be shared and how. (RLG-OCLC, 2002). It is also important to make

a plan for preservation strategies and shared responsibilities for executing them in a manner that is sustainable over the long term. Creation and implementation of preservation metadata is another task required for digital preservation. Automated capturing of the metadata has been examined as a cost-effective approach to the development of models and software for digital preservation (Hitchcock et al., 2005). Along with this, IR staff responsible for metadata services needs to have knowledge about the characteristics, scope, file formats, and subject matter of IR collections, as well as the context in which the collections were created, used, and modified. The desire of faculty to self-archive and contribute data as well as the published end product should be taken quite seriously by IR staff. Many of the current lists of file format that will be sustained for the long term include those common for end products, such as Word or PDF but not those associated with data.

Concerning copyright management, it is essential that IRs build ongoing relationships with other stakeholders in the issue on campus and in the scholarly communication process. These stakeholders include not only publishers and faculty but university presses, scholarly publishing office personnel, and legal counsel. Developing guidelines of copyright management for self-archiving helps faculty and students become more aware of the issues and alleviates their concerns. A good example of operating services for copyright management is the University of Connecticut Library Copyright Team (Oakley et al., 2007). This team developed a web site containing general copyright information, and links to related library services, such as the IR. Collaboration with various stakeholders is pivotal in such a project and this requires librarians' ability to identify 'who knows what' and to maintain working relationships with them.

Limitations

The present study has limitations with regard to its internal and external validity. In order to test the internal validity, I performed a confirmatory factor analysis, which verifies "a particular pattern of relationships predicted on the basis of theory or previous analytic results" (DeVellis, 2003, p.131). The factor analysis shows that my model of factors affecting self-archiving has a good fit to the survey data based on RMSEA (Root Mean Square Error of Approximation), which is close to 0.06 (see Appendix D). However, it is possible that there is more than one model that accurately describes the data, and thus there is no evidence that my model is optimal. This indicates that an alternative structure of factors influencing self-archiving could be developed and it may also provide additional plausible relationships among factors.

Concerning external validity, survey respondents may be more interested in self-archiving practices than non-respondents, and in fact, the majority of respondents self-archived already. This pattern of self-archiving may not be generalizable to the entire population of faculty in research universities. In addition, I often grouped survey respondents by using the disciplinary categories of Science, Engineering, Social Science and Humanities, when analyzing disciplinary differences in self-archiving behavior. Some disciplines, however, did not fit smoothly in these four areas. It is also possible that different perceptions or behavior of self-archiving within each area can exist but these were not examined in detail. This makes it difficult to generalize self-archiving behavior of faculty from each group of disciplines.

Directions for Future Research

This study identifies several factors that either positive or negative relationships with self-archiving behavior measured as two dependent variables. One concerns self-archiving research works on any kinds of publicly accessible web sites. The other relates

to self-archiving in university IRs. IR contribution was one type of the general self-archiving behavior, although only two factors were commonly found to be significant - copyright concerns and altruism. Nonetheless, the factor of copyright concerns was negatively associated with general self-archiving behavior, whereas it was positively related to IR contribution. This result indicates that factors affecting self-archiving vary among different types of venues. This study did not examine factors affecting self-archiving in any specific venue except for IRs, and there is room for future research on the topic. For example, those who have contributed to disciplinary repositories may be influenced significantly by self-archiving culture, since such repositories emerged in disciplines with a tradition of sharing pre-prints. Understanding these factors would help evaluate the usefulness of a particular self-archiving venue and make suggestions for the improvement of its functions and services.

In addition, the majority of faculty participants in this study have employed multiple venues for self-archiving. Only 60 (12.5%) out of 480 faculty self-archivers utilized a single venue. This raises a question about how to characterize self-archivers of a particular venue. In this study, real IR contributors were defined to be respondents who self-reported that they have contributed to the IRs. However, they can be self-archivers in personal web sites or disciplinary repositories. Future research needs to consider additional dimensions that help determine self-archivers in a specific venue, for instance, the extent to which faculty members self-archive in a certain venue, e.g., frequency of self-archiving or the number of research contents self-archived.

Moreover, the updated model (Figure 5.1) delineates relationships between various factors and faculty self-archiving behavior. The OLS regression analysis shows standard coefficient (β) values, which indicate the effect of factors on self-archiving (Table 4.30). Based on the β values, the factor of altruism is the strongest factor (β=0.28), followed by self-archiving culture (β=0.27). The remaining significant factors have β values ranging from 0.09 to 0.15. However, some of those factors appear correlated with

one another, for example rank and age, and self-archiving culture and technical skills. Although the OLS regression model showed no multicollinaearity problem, those seemingly correlated factors need to be examined more. For instance, rank and age were positively associated with one another. Yet, rank has a positive relationship with self-archiving, while age was negatively related with it. According to β values, rank has slightly greater value (β=0.12) than age does (β=-0.10), but it is unclear whether this difference is meaningful to conclude that rank has greater impact on self-archiving than age. Future research need to clarify the correlation among independent variables and the degree of impact that each variable has upon faculty self-archiving. Clustering independent variables into principal dimensions would help identify important factors and make the interpretation straightforward.

In conclusion, this study demonstrates several factors associated with faculty's decision to self-archive. Especially, the factor of altruism was found to be the greatest effect on general self-archiving behavior and the second largest impact on IR contribution. Based on interview data analysis, this altruism was driven by the idea of generalized reciprocity. Faculty members benefited from OA content were motivated to self-archive their research with hope to return the benefit to whoever using it. As a result, faculty members were encouraged to self-archive by the intrinsic benefit, rather than extrinsic benefits, such as increased citation rates and impact of OA research. In addition, the study suggests two aspects of IR practices that would increase faculty participation – digital preservation and copyright management. Collaboration with various stakeholders is essential for both these activities. IRs should assess needs, priorities, and resources for digital preservation in their organizational context, and comply with best practices and standards that are emerging nationally (CRL-OCLC, 2007). IRs could also coordinate information resources and expertise in copyright available on campus to notify faculty how to manage copyright in their self-archiving practices. Guidelines for copyright

management would be helpful for faculty and students who consider IR contribution and minimize their concerns about the issue.

Self-archiving represents significant changes in the dissemination of research, in terms of its accessibility, timeliness, and cost-effectiveness. Despite these benefits, faculty participants in this study emphasized the value of existing scholarly communication system and considered self-archiving to be supplementary. Self-archiving practices reside in a broad socio-technical network involved in scholarly communication. The study delineated interactions among social and technical factors and attempted to answer the reasons of low faculty participation in certain self-archiving practices, particularly IRs. This study implies that IR staff need to establish appropriate service models that satisfy contributors' needs and promote faculty participation.

APPENDICES

Appendix A. Final Survey Instrument

Consent Form

The purpose of this survey is to understand university faculty members' motivations for and perceived barriers to posting research/teaching materials on publicly accessible web sites. This activity is termed "self-archiving", which represents changes in ways of disseminating scholarly content through the Internet. Since faculty members are primary authors of scholarly content, understanding your perceptions on self-archiving and methods that you use for the activity, if any, are critical.

This study will contribute to research on the transformation of scholarly publishing paradigm resulting from self-archiving through such methods as personal homepages, project web sites, and open access repositories. In addition, findings of this survey will benefit repositories for self-archiving, such as Institutional Repositories (IRs), to design and implement better services for contributors and users.

Survey participants will not face any risks throughout the study. Although research results may be published, you will not be identified in any reports on this study. All research records will be kept confidential to the extent provided by federal, state, and local laws. Data from the study will be retained in a password-protected and secure server, as well as in CD-ROMs, which will be kept under lock and key. The data will have identifier-free codes and not identified according to personal names. This consent form will be kept together with the research records on this study.

Your participation in this project is voluntary. Subsequent to your consent, you may refuse to answer specific questions, participate in, or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled. You may also ask questions concerning the study, before, during, or after the survey.

Should you have questions concerning your rights as a research participant, you should contact the Institutional Review Board, 540 East Liberty Street, Suite 202, Ann Arbor, MI 48104-2210, (734) 936-0933, irbhsbs@umich.edu

For all other questions, please contact the principle investigator:

Jihyun Kim
Ph.D. Candidate
School of Information
University of Michigan
1075 Beal Ave, Room xxxx
Ann Arbor, MI 48109-2112
Voice: (734) xxx-xxxx
Fax: (734) xxx-xxxx
Email: jhkz@umich.edu

The survey will take approximately 15 minutes. Thank you for your participation.

I have read the information in this consent form, and I agree to participate in this study.

O I agree
O I do not agree

Page 1 of 10

1. Have you ever made your research/teaching materials publicly accessible on the Internet?

   O Yes
   O No

Page 2 of 10

Questions 2-a to 2-f ask the frequency of posting several types of research materials in various publicly accessible web sites.

*** Definition ***

A **subject-based open archive refers to a publicly accessible and field-wide repository where researchers post their pre- or post-refereed research articles. (e.g., ArXiv in Physics, MPRESS in Mathematics, or CogPrints in Psychology)*

*** An **Institutional Repository (IR)** is a set of services and technologies to collect, disseminate and preserve research and teaching materials in digital form that were created by university faculty, students and staff. The materials in the IR are publicly accessible through the Internet. Your university currently implements the IR.*

2-a. In the past 3 years, how many times have you deposited ***pre-refereed drafts*** in the following ways?

	None	Once	2-3 times	4-5 times	6+ times
Personal web pages	O	O	O	O	O
Research group/lab/center web sites	O	O	O	O	O
Department/school/college web sites	O	O	O	O	O
Subject-based open archives*	O	O	O	O	O
Institutional Repository**	O	O	O	O	O
Other	O	O	O	O	O

2-b. In the past 3 years, how many times have you deposited ***publishers' PDF files of refereed articles*** in the following ways?

	None	Once	2-3 times	4-5 times	6+ times
	O	O	O	O	O

Personal web pages					
Research group/lab/center web sites	O	O	O	O	O
Department/school/college web sites	O	O	O	O	O
Subject-based open archives*	O	O	O	O	O
Institutional Repository**	O	O	O	O	O
Other	O	O	O	O	O

2-c. In the past 3 years, how many times have you deposited your ***final versions of refereed articles*** in the following ways?

	None	Once	2-3 times	4-5 times	6+ times
Personal web pages	O	O	O	O	O
Research group/lab/center web sites	O	O	O	O	O
Department/school/college web sites	O	O	O	O	O
Subject-based open archives*	O	O	O	O	O
Institutional Repository**	O	O	O	O	O
Other	O	O	O	O	O

2-d. In the past 3 years, how many times have you deposited ***unrefereed articles*** (technical reports, working papers, or project reports) in the following ways?

	None	Once	2-3 times	4-5 times	6+ times
Personal web pages	O	O	O	O	O
Research group/lab/center web sites	O	O	O	O	O
Department/school/college web sites	O	O	O	O	O
Subject-based open archives*	O	O	O	O	O
Institutional Repository**	O	O	O	O	O
Other	O	O	O	O	O

2-f. In the past 3 years, how many times have you deposited ***book chapters*** in the following ways?

	None	Once	2-3 times	4-5 times	6+ times
Personal web pages	O	O	O	O	O
Research group/lab/center web sites	O	O	O	O	O
Department/school/college web sites	O	O	O	O	O
Subject-based open archives*	O	O	O	O	O
Institutional Repository**	O	O	O	O	O
Other	O	O	O	O	O

3. If you have used other types of publicly accessible web sites to post any of those materials above, please state them.

4. Who was involved in depositing your work into any of the above? (Select all that apply)

- O Myself
- O Students or assistants
- O Department staff
- O Library staff
- O Collaborators
- O Other (Please specify)

5. How long have you been depositing papers in any of the above?

- O 1 year or less
- O 1-3 years
- O 4-5 years
- O More than 5 years

6. What percentage of following your work produced in the last 5 years have you made publicly accessible on the Internet?

	None	1-25%	26-50%	51-75%	76-100%	I don't know
Pre-refereed draft	O	O	O	O	O	O
Refereed, published articles	O	O	O	O	O	O
Unrefereed articles (technical reports, working papers, or project reports)	O	O	O	O	O	O
Book chapters	O	O	O	O	O	O
Data sets	O	O	O	O	O	O

Page 3 of 10

7. What other types of materials have you made publicly accessible on the Internet? (Select all that apply)

O Dissertation or thesis
O Conference presentations
O Software
O Software documentation
O Images
O Audio/video recordings
O Course syllabi
O Lecture notes, handouts and assignments
O NONE of these
O Other (please specify)

8. Are you aware of the Creative Commons?

O Yes
O No

9. How many times have you used the Creative Commons license?

O Never
O Once
O 2-3 times
O 4-5 times
O More than 5 times

Page 4 of 10

An Institutional Repository (IR) is a set of services and technologies to collect, disseminate and preserve research and teaching materials in digital form that were created by university faculty, students and staff. The materials in the IR are open and publicly accessible through the Internet.

10. Are you aware of the Institutional Repository (IR) in your university?

O Yes
O No

Page 5 of 10

11. How did you learn about the Institutional Repository (IR) in your university? (Select all that apply)

- O IR staff contacted me.
- O IR staff made a presentation at departmental and/or faculty meetings.
- O The Dean or chair of my school / department told faculty about the IR.
- O Another university authority told me about the IR.
- O Faculty colleagues told me about the IR.
- O I found publicity about the IR in university / library web sites
- O I read publicity about the IR in campus newspapers
- O Other (please specify)

12. Have you searched the IR at your institution?

- O Yes
- O No

Page 6 of 10

Page 6 and 7 provides a series of statements regarding making your research and/or teaching materials publicly accessible through the Internet..

13-a. Please indicate to what extent you agree or disagree with the following statements.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
I need to ask permission from publishers to post my work on publicly accessible web sites.	O	O	O	O	O	O
If I post my work on publicly accessible web sites without permission, I may infringe on copyright.	O	O	O	O	O	O
I need permission from co-authors or collaborators to post my work on publicly accessible web sites.	O	O	O	O	O	O
I cannot publish my work if I post it on publicly accessible web sites before publication.	O	O	O	O	O	O

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
Posting my materials on publicly accessible web sites takes time away from my research and writing.	O	O	O	O	O	O
Additional time and effort is required to make my materials publicly accessible on the	O	O	O	O	O	O

Internet.						
It is difficult to learn how to enter the required data (e.g., title, author, date etc.) with my deposit.	O	O	O	O	O	O
Posting my research work on publicly accessible web sites will increase the chance to communicate my research findings to peers.	O	O	O	O	O	O
Materials on publicly accessible web sites are more easily accessible through Internet search engines.	O	O	O	O	O	O
Materials on publicly accessible web sites are not preserved in perpetuity.	O	O	O	O	O	O
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
Posting my materials on publicly accessible web sites will enlarge the readership of the materials.	O	O	O	O	O	O
Posting my research work on publicly accessible web sites will increase the potential impact of my work.	O	O	O	O	O	O
Posting my research work on publicly accessible web sites allows for earlier dissemination of my research findings.	O	O	O	O	O	O
Posting my research work on publicly accessible web sites will increase my visibility within the discipline(s) to which I belong.	O	O	O	O	O	O
Materials on publicly accessible web sites will be cited more frequently.	O	O	O	O	O	O
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
My university will accept research work on publicly accessible web sites as an alternative to publication for tenure and promotion.	O	O	O	O	O	O
Posting my work on publicly accessible web sites will adversely affect my chances of tenure/promotion.	O	O	O	O	O	O
Posting my research work on publicly accessible web sites will adversely affect my chances of attaining research grants.	O	O	O	O	O	O

Page 7 of 10

14-a. Please indicate to what extent you agree or disagree with the following statements.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
I will continue posting my work on publicly accessible web sites even if others in my field do not.	O	O	O	O	O	O
I support the principle of open access (free and unrestricted access to research materials) for all users.	O	O	O	O	O	O
Posting my materials on publicly accessible web sites will help other researchers build on my research findings.	O	O	O	O	O	O
Posting my materials on publicly accessible web sites allows other scholars to access those that they could not otherwise use.	O	O	O	O	O	O

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
I trust the quality of materials on publicly accessible web sites from authors employed by prestigious institutions.	O	O	O	O	O	O
I trust the quality of materials on publicly accessible web sites from well-known researchers in my field.	O	O	O	O	O	O
I trust the quality of peer-reviewed articles on publicly accessible web sites.	O	O	O	O	O	O
If I post my materials on publicly accessible web sites, readers may plagiarize or fail to cite my work.	O	O	O	O	O	O
If I post my materials on publicly accessible web sites, the integrity of my work will be compromised.	O	O	O	O	O	O
Materials on publicly accessible web sites are not maintained securely.	O	O	O	O	O	O

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
In my field, it is common for researchers to post their pre- or post-refereed articles on publicly accessible web sites.	O	O	O	O	O	O

In my department, it is common for faculty and students to create and share working papers or technical reports.	O	O	O	O	O	O
My decision to make, (or not to make) my materials publicly accessible on the Internet was influenced by my co-authors or collaborators.	O	O	O	O	O	O
My decision to make, (or not to make) my materials publicly accessible on the Internet was influenced by my grant-awarding body.	O	O	O	O	O	O
My decision to make, (or not to make) my materials publicly accessible on the Internet was influenced by my university or department.	O	O	O	O	O	O
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	I don't know
I am concerned about the cost to my university of subscription-based journals.	O	O	O	O	O	O
I feel a sense of belonging to my university.	O	O	O	O	O	O
I am proud to be a professor in my university.	O	O	O	O	O	O

15. Do you have other opinions or concerns regarding faculty self-archiving practices? Please state them.

Page 8 of 10

16. Do you plan to post your research/teaching materials on publicly accessible web sites in the future?

O Yes
O No

O Uncertain

Page 9 of 10

17. What types of publicly accessible web sites do you plan to use in the future? (Select all that apply)

O Personal web pages
O Research group/lab/center web sites
O Department/school/lab web sites
O Subject-based open archives (publicly accessible and field-wide repositories)
O Institutional Repository (IR)

O Other (please specify)

18. If you plan to contribute to the IR in the future, how important are these reasons for you to contribute to the IR?

*** If you do not plan to contribute to the IR, skip this and the next questions and go to the next page.***

	Very Important	Somewhat Important	Neutral	Somewhat unimportant	Very unimportant
My contribution would count toward my tenure and promotion.	O	O	O	O	O
Publishers would not have exclusive rights over my work.	O	O	O	O	O
If there were a peer review process in the IR	O	O	O	O	O
The IR would preserve my materials.	O	O	O	O	O

I would receive financial reward.	O	O	O	O	O
If the IR shows how many times my materials in the IR were viewed and downloaded	O	O	O	O	O
I would receive recognition from my university.	O	O	O	O	O

19. Do you have other reasons that will motivate you to contribute to the IR? Please describe them.

Page 10 of 10

Demographic

20. Which of the following best describes your current professional rank?

- O Assistant professor
- O Associate professor
- O Full professor
- O Distinguished professor
- O Others (please specify)

21. Which of the following roles, if any, have you undertaken in the past year? (Select all that apply)

- O Authors of journal articles
- O Referee for journal articles
- O Society editorial board members
- O Journal editorial board members
- O Senior or Managing journal editor
- O Other (please specify)

22. Which of the following positions, if any, have you ever held? (Select all that apply)

- O Tenure and promotion committee member
- O Department chair
- O Director of research center/lab
- O Dean/Associate dean
- O Provost
- O Other (please specify)

23. Approximately how many articles do you publish each year?

O 0-1
O 2-3
O 4-5
O More than 5

24. If any of your public and private grant funders take a position on making your research data publicly accessible, the grant funder(s) _____ (Choose one)

O Mandate deposit.
O Encourage deposit.
O Neither encourage nor discourage deposit.
O Discourage deposit.
O Other (please specify)

25. In reference to question 24, please state the grant-awarding body, which most affected your response.

26. I have worked at my university as a professor since (please provide year)

27. In what year were you granted your doctoral degree?

28. Rate your ability to do each of the following (Select the appropriate number: 1= no knowledge; to 5=expert user).

	1	2	3	4	5
Use a spreadsheet or database program on a computer	O	O	O	O	O
Send and receive e-mail	O	O	O	O	O
Search for information on the Internet/World Wide Web	O	O	O	O	O
Create or edit a World Wide Web site	O	O	O	O	O
Program a computer using a programming language (such as C, C++, Java)	O	O	O	O	O

29. Year of birth

30. Gender
O Male
O Female

Appendix B. Invitation Letter

October 27, 2006

Dear Professor XXX:

You have been selected to participate in a survey entitled "Faculty Self-Archiving Behavior". We would greatly appreciate your taking 15 minutes to respond.

> Definition: Self-archiving refers to depositing research and/or teaching materials on publicly accessible web sites, such as personal homepages, research project web sites, or open access repositories.

This study is funded by the Institute of Museum and Library Services (IMLS) and is being administered to faculty in select colleges and universities around the United States as part of a larger investigation into Institutional Repositories nationwide. A goal of the research is to better understand university faculty members' motivations for and perceived barriers to self-archiving their research and/or teaching materials. Self-archiving activities represent a major change in the way faculty disseminate scholarly content. Therefore, your perceptions on self-archiving and methods that you use for this activity are critical.

As a token of our appreciation for your efforts, please accept the enclosed cash gift. Soon, we will be sending you an e-mail with a link to the online survey. You can also take the survey now by going to **«URL»** and entering **Survey ID # XXXX**. If possible, please complete the online questionnaire by November 17.

Your participation is voluntary. You may skip questions that make you uncomfortable and are free to withdraw at any point. Your responses will be used for research purposes only and will be kept in secure locations at the University of Michigan School of Information. The information you provide in this survey will be kept confidential and responses will only be reported in aggregate form. Furthermore, all personal information will be presented only in an aggregate form in reports and publications. Individual responses will not be identifiable. If you have any questions regarding your rights as a participant in this research, please contact: Institutional Review Board, 540 East Liberty Street, Suite 202, Ann Arbor, MI 48104-2210, Tel: (734) 936-0933, E-mail:<irbhsbs@umich.edu>.

Thank you in advance for taking the time to complete this important questionnaire. More information about this and related studies is available online at: http://miracle.si.umich.edu. If you have additional questions or concerns, please contact us via email at jhkz@umich.edu or by calling 734-647-8041.

Sincerely,

Elizabeth Yakel Jihyun Kim

Appendix C. Interview Protocol

Interview Protocol for Self-archivers

- Tell me about your last experience posting your work on a publicly accessible web site.
- Tell me about your decision-making process in making your work publicly accessible on the Internet.
    - How do you decide what types of research/teaching work you post on the Internet in which web sites?
- Why do you make certain materials publicly accessible?
- Are there materials you prefer to keep private?
- Why do you use certain publicly accessible web sites for posting your work?
    - Who actually post your work on the web site?
- What are your thoughts about posting your work on publicly accessible web sites?
    - What do you think about copyright issues?
        - How aware are you of publishers' policies relating to posting your research work on publicly accessible forums?
    - What do you think about additional time and effort required for posting your work?
    - What do you think about the influence of this behavior on tenure and promotion?
    - What is your thought about long-term accessibility of your work publicly accessible on the web?
        - How do you maintain your work-related files on your computer?
    - Have you thought that making your work publicly accessible on the web helps other researchers build on research, or allow others to access your work that they could not otherwise?
    - How common is this behavior in your field (or department)?
    - Do you have concerns about other readers using your work that are publicly accessible on the web?
- What do you think are the benefits of posting your work on publicly accessible web sites?
    - Are there benefits to early dissemination in your field?
    - By making your work publicly accessible on the web, have you experienced increased communication to peers?
    - What do you thing about citation rate of your work publicly accessible online?
    - What do you think about impact/recognition of your work?
    - Others?
    - How do you know there are such benefits of posting your work on the Internet?
- Do your grant-awarding bodies influence your decision to self-archive?
- How much do your collaborators or peers in your field affect your decision to self-archive?

- How much does your university/department affect your decision to self-archive?
- Do you contribute to the Institutional Repository (IR) in your university?
- If yes, what does motivate you to deposit your work? If no, what would motivate you to contribute?

Interview Protocol for non-self-archivers

- [Only for potential self-archivers] Based on your survey answers, you plan to self-archive your work in publicly accessible web sites. What types of your work would you be willing to make publicly accessible on the Internet?
- [Only for potential self-archivers] What do you think are the potential benefits of posting your work on publicly accessible web sites?
- What are your thoughts about posting research/teaching work on publicly accessible web sites?
    - What do you think about copyright issues?
        - How aware are you of publishers' policies relating to self-archiving your posting your research work on publicly accessible forums?
    - What is your thought about your time and effort required for posting your work?
    - Tenure and promotion
    - Do you have concerns about other readers using your work if you make it publicly available on the web?
    - What do you think about the long-term accessibility of research/teaching materials publicly available on the web?
    - How common is this behavior in your field (or department)?
- Are there (research/teaching) materials you prefer to keep private?
- What makes you reluctant to posting your work on publicly available web sites?
- [Only for non-self-archivers] Do you see some benefits posting your work on publicly available web sites?
- How much do your collaborators or peer in your field affect your decision not to self-archive?
- Do your grant-awarding bodies have any preferences for your self-archiving?
- Does your university/department have any preferences for your self-archiving?
- Do you plan to contribute to the Institutional Repository (IR) in your university in the future?
    - If yes, what does motivate you to deposit your work? If no, what would motivate you to contribute?

Appendix D. Statistical Methods

Evaluating the models

The present study developed two statistical model – (1) the OLS regression model (Table 4.30) that examined factors affecting self-archiving; (2) the logistic regression model (Table 4.33) that determined factors influencing IR contribution. In order to make valid statistical inferences, I assessed the models to see whether the survey data met the assumptions underlying OLS and logistic regression.

Normality and Homoscedasticity of Residuals

These assumptions were related to the OLS regression model. The normality assumption indicated that the errors should be normally distributed. To check normality of residuals, a normal probability plot was generated that plotted against theoretical normal distribution. The plot showed an approximate straight line, which suggested that the residuals of the OLS model were normally distributed. The other assumption – homoscedasticity of residuals indicated that error variance should be constant. I performed Szroeter's test for homoscedasticity, of which the null hypothesis was that residual variance was constant. The p-value generated from the test was .1607, which meant that the null hypothesis failed to be rejected. Therefore, the OLS model satisfied the homogeneity of residual variance.

The Logistic Regression Model Adequacy

The logistic regression model was based on the assumption that the logit of the dependent variable was a linear combination of independent variables. In order to assess the overall specification of the model, I performed the link test, which computed a new regression model using the score of independent variables from the original model and its square as predictor variables. If the squared value was significant, there was evidence that the model was not correctly specified. The result of the link test showed that the squared

value was not significant (p=0.537) and therefore, no model specification error was detected. In addition, I performed the Hosmer and Lemeshow's goodness-of-fit test, which created the Pearson chi-square and a contingency table of observed and expected frequencies. A good fit measured by the test generated a large p-value. The resulting p-value was .9084, which suggested that the logistic regression model fitted the survey data well.

Multicollinearity

Multicollinearity in a multiple regression model indicated that two or more independent variables are highly correlated. This problem can be detected by calculating Variance Inflation Factor (VIF), although there was no formal value of VIF to determine the existence of multicollinearity. The rule of thumb is that there would be high mulicollinearity if a maximum VIF value exceeds 10, and a mean of VIF value is significantly greater than 1. In this study, the maximum VIF value was 2.51 and the mean VIF was 1.68, and thus mulicollinearity was not a problem in both OLS and logistic regression models.

Removing potentially troublesome items

I decided to exclude three items (item#3, 10, and 19 in Table 3.3), which reduced the reliability (Cronbach's alpha) of their constructs. By removing these items, the alphas of relevant dimensions – copyright, accessibility, and academic reward - became greater than .70, or close to the acceptable level (Table 4.24).

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) is employed to test hypotheses concerning the factor structure predetermined based on literature. By using this method, researchers are able to match observed and theoretical factor structure for a given data set to identify goodness-of-fit of the predetermined factor model. In the present study, CFA was performed using Amos7. The value of RMSEA (Root Mean Square Error of Approximation) was .06, the cutoff criterion recommended by Hu and Bentley (1999). In

addition, R-square statistics showed how much variance the respective factor accounted for in observed scale questions. The R squares ranged from .379 to .882, which indicated a respectable portion of variance explained. As a result, the factor structure examined in this study showed reasonably good model fit.

Appendix E. Disciplines of Survey Respondents

Disciplines of Respondents from Science and Engineering fields

	Science	Freq.	Engineering	Freq.
IR contributors	Mathematics	17	Electrical Engineering & Computer Science	32
	Physics (& Astronomy)	4	Computer Science	25
	Chemistry	2	Mechanical Engineering	19
	Biology	2	Horticultural Science	5
	Ecology & Evolutionary Biology	2	Chemical Engineering	4
	Geology	2	Aerospace Engineering	4
	Medicinal Chemistry	1	Civil and Environmental Engineering	3
	Zoology	1	Food Science	3
	Health Informatics	1	Materials Science and Engineering	3
	Physical Therapy	1	Bioresources Engineering	2
	Pharmacy	1	Earth and Planetary Sciences	2
			Natural Resources	2
			Marine and Earth Studies	1
			Animal and Avian Sciences	1
			Architecture	1
			Crop and Soil Sciences	1
			Design and Environmental Analysis	1
			Environmental Science and Technology	1
			Nuclear Science and Engineering	1
			Industrial & Systems Engineering	1
			Polymer, Textile & Fiber Engineering	1
			Plant Pathology	1
	Total	34	**Total**	114
	Science	**Freq.**	**Engineering**	**Freq.**
IR non-contributors	Physics (& Astronomy)	44	Computer Science	20
	Biology (Molecular or Cell Biology)	37	Electrical Engineering & Computer Science	34
	Mathematics	46	Mechanical Engineering	24
	Total	127	**Total**	78
	Grand Total	161	**Grand Total**	192

Disciplines of Respondents from Social Science and Humanities

	Social Science	Freq.	Humanities	Freq.
IR contributors	Economics	27	Linguistics	8
	Business	25	History	3
	Psychology	9	German	2
	Education	9	Anthropology	1
	Sociology	6	Philosophy	2
	Public Policy	5	English	1
	Political Science	4	French	1
	Information Science	4	Art and Art History	1
	Brain and Cognitive Science	2	Latin American Studies	1
	Social Welfare	1	Romance Languages	1
	Applied Behavioral Science	1	Slavic Languages & Literatures	1
	Public and Community Health	1	Speech-Language-Hearing	1
	Urban Studies and Planning	1	Literature, Communication, and Culture	2
	Social statistics	1		
	Total	96	**Total**	25
	Social Science	**Freq.**	**Humanities**	**Freq.**
IR non-contributors	Economics	27	Art (& Art History)	24
	Psychology	41	English (or literature)	40
	Sociology	42	History	36
	Total	110	**Total**	100
	Grand Total	206	**Grand Total**	125

BIBLIOGRAPHY

Allen, J. (2005). *Interdisciplinary differences in attitudes towards deposit in institutional repositories.* Master's thesis, Manchester Metropolitan University, Manchester, UK. http://eprints.rclis.org/archive/00005180/01/FULLTEXT.pdf

Andrew, T. (2003). Trends in self-posting of research material online by academic staff. *Ariadne, 37.* http://www.ariadne.ac.uk/issue37/andrew/

Antelman, K. (2004). Do Open-access articles have a greater research impact? *College & Research Libraries, 65*(5), 372-382.

Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing, 19*(2), 85-95.

arXiv.org. (2006). arXiv.org e-Print archive. http://www.arxiv.org/

Birnholtz, J., D. Horn, T. Finholt, & S. Bae. (2004). The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of a technologically sophisticated sample. *Social Science Computer Review, 22*(3). http://kmdi.utoronto.ca/jeremy/SSCORE.pdf

Bishop, A. P. (1995). Scholarly journals on the net: A reader's assessment. *Library Trends, 43*, 544-570.

Björk, B. (2004). Open access to scientific publications - an analysis of the barriers to change? *Information Research, 9*(2). http://informationr.net/ir/9-2/paper170.html

BOAI. (2006). Budapest Open Access Initiative: Frequently Asked Questions. http://www.earlham.edu/~peters/fos/boaifaq.htm

Bohin, I. (2004). Communication regimes in competition: The current transition in Scholarly Communication seen through the lens of the Sociology of Technology. *Social Studies of Science*, 34(3), 364–391.

Borgman, C. L. (2000). *From Gutenberg to the global information infrastructure: access to information in the networked world*. Cambridge, MA: MIT Press.

Borgman, C. L. (2003). Designing digital libraries for usability. In A. P. Bishop, N. A. Van House & B. P. Buttenfield (Eds.), *Digital library use: social practice in design and evaluation*. Cambridge, MA: MIT press.

Branin, J. (2005). Institutional repositories. In M. A. Drake (Ed.), *Encyclopedia of Library and Information Science* (pp. 237-248). Boca Raton, FL: Taylor & Francis Group, LLC. http://hdl.handle.net/1811/441

Brennan, M. J., J. M. Hurd, D. D. Blecic, & A. C. Weller. (2002). A Snapshot of early adopters of e-Journals: challenges to the library. *College & Research Libraries, 63,* 515-526.

Brown, C. (1999). Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians and physicists. *Journal of the American Society for Information Science, 50,* 929-943.

Brown, C. (2001). The E-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science and Technology, 52*(3), 187-200.

Brown, C. (2003). The role of electronic preprints in chemical communication: analysis of citation, usage, and acceptance in the journal literature. *Journal of the American Society for Information Science and Technology, 54*(5), 362-371.

Carr, L., & S. Harnad. (2005). Keystroke economy: A study of the time and effort involved in self-archiving. http://eprints.ecs.soton.ac.uk/10688/01/KeystrokeCosting-publicdraft1.pdf

Chan, L. (2004). Supporting and enhancing scholarship in the digital age: the role of open-access institutional repositories. *Canadian Journal of Communication, 29,* 277-300.

Coleman, A. (2006). Self-Archiving and the Copyright Transfer Agreements of ISI-Ranked Library and Information Science Journals. http://dlist.sir.arizona.edu/1041/

Coleman, A., & J. Roback. (2005). Open Access Federation for Library and Information Science: dLIST and DL-Harvest. *D-Lib Magazine, 11*(12). http://www.dlib.org/dlib/december05/coleman/12coleman.html

Constant, D., S. Kiesler, & L. Sproull. (1994). What's mine is ours, or is it? A study of attitude about information sharing. *Information Systems Research, 5*(4), 400-421.

CRL-OCLC (2007). Trustworthy Repositories Audit & Certification (TRAC): Criteria and checklist. http://www.crl.edu/PDF/trac.pdf

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

Cropanzano, R. & M. S. Mitchell (2005). Social Exchange Theory: An interdisciplinary review. *Journal of Management*, 31(6), 874-900.

Crow, R. (2002). The case for institutional repositories: A SPARC position paper. http://www.arl.org/sparc/IR/ir.html

Crow, R., & H. Goldstein. (2004). *Guide to business planning for converting a subscription-based journal to open access*: Open Society Institute. http://www.soros.org/openaccess/oajguides/business_converting.pdf

Davis, P. M. & M. J. L. Connolly. (2007). Institutional Repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*: 13(3/4) http://www.dlib.org/dlib/march07/davis/03davis.html

Davis, P. M., & J. Fromerth (2007). Does the ArXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics, 71(2): 203-215.

DeVellis, R. F. (2003). *Scale development: theory and applications*. Thousand Oaks, CA: Sage Publications.

Digital Preservation Management Tutorial (2003). Digital preservation management: Implementing short-term strategies for long-term problem. http://www.icpsr.umich.edu/dpm/dpm-eng/contents.html

Foster, N. F., & S. Gibbons. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine, 11*(1). http://www.dlib.org/dlib/january05/foster/01foster.html

Fowler, F. J. J. (1995). *Improving survey questions: design and evaluation*. Thousand Oaks, CA: Sage publications.

Gadd, E., C. Oppenheim, & S. Probets. (2003a). RoMEO studies 1: The impact of copyright ownership on academic author self-archiving. *Journal of Documentation, 59*(3), 243-277.

Gadd, E., C. Oppenheim, & S. Probets. (2003b). RoMEO studies 2: How academics want to protect their open-access research papers. *Journal of Information Science, 29*(3), 333-356.

Garvey, W. D. (1979). *Communication: the essence of science*. New York, NY: Pergamon Press.

Garvey, W. D., & B.C. Griffith. (1972). Communication and information processing within scientific disciplines: empirical findings for psychology. *Information and Storage, 8*, 123-126.

Garvey, W. D., & B. C. Griffith. (1967). Scientific communication as a social system. *Science, 157*(3792), 1011-1016.

Ginsparg, P. (2002). Can peer review be better focused? *Science and Technology Libraries*, 22(3-4), 5-17.

Ginsparg, P. (2007). Next generation implications of Open Access. *CTWatch Quarterly*, 3(3): 11-18. http://www.ctwatch.org/quarterly/pdf/ctwatchquarterly-12.pdf

Gomes, S., & J. Meadows. (1998). Perceptions of electronic journals in British universities. *Journal of Scholarly Publishing, 29*, 174-181.

Goodman, D., S. Dawson, & J. Yaremchuk. (2007). Open access and accuracy: A comparison of authors' self-archived manuscripts and published articles. *Learned Publishing*, 20(3), 203-215.

Greenberg, J., M. C. Pattuelli, B. Parsia, & W. D. Robertson. (2002). Author-generated Dublin Core metadata for web resources: A baseline study in an organization. *Journal of Digital Information, 2*(2). http://ojfpc.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/

Guédon, J.-C. (2003). Open access archives: from scientific plutocracy to the republic of science. *IFLA Journal, 29*(2), 129-139.

Guédon, J.-C. (2004). The "green" and "gold" roads to open access: the case for mixing and matching. *Serials Review, 30*(4), 315-328.

Gunnarsdóttir, K. (2005). Scientific journal publication: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science*, 35(4): 549–579.

Hall, H. (2001, April 10-11, 2001.). *Social exchange for knowledge exchange.* Paper presented at the Managing Knowledge: Conversations and Critiques, University of Leicester Management Centre. http://www.soc.napier.ac.uk/publication/op/getpublication/publicationid/321908

Hall, H. (2003). Borrowed theory: applying exchange theories in information science research. *Library and Information Science Research, 25*(3), 287-306.

Harley, D. (2008). The University as Publisher: Summary of a Meeting Held at UC Berkeley on November 1, 2007. *Journal of Electronic Publishing*, 11(2). http://hdl.handle.net/2027/spo.3336451.0011.208

Harnad, S. (2003a). Online archives for peer-reviewed journal publications. In J. Feather & P. Sturges (Eds.), *International Encyclopedia of Library and Information Science*. London: Routledge. http://cogprints.org/3020/01/eprints.htm

Harnad, S. (2003b). Open Access to peer-reviewed research through author/institution self-archiving: Maximizing research impact by maximizing online access. *Journal of Postgraduate Medicine*, 49(4), 337-342. http://www.jpgmonline.com/text.asp?2003/49/4/337/4978

Harnad, S. (2004). *Re: Free Access vs. Open Access*. Retrieved July 7, 2006, from https://arl.org/Lists/SPARC-IR/Message/192.html

Harnad, S., & T. Brody. (2004). Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine, 10*(6). http://www.dlib.org/dlib/june04/harnad/06harnad.html#Pringle

Harnad, S. (2006) Opening access by overcoming Zeno's paralysis. In Jacobs, N., (Ed) *Open Access: Key Strategic, Technical and Economic Aspects*. Chandos Publishing (Oxford) Limited, Chapter 8.

Harter, S. P. (1998). Scholarly communication and electronic journals: an impact study. *Journal of the American Society for Information Science, 49*(6), 507-516.

Hedstrom, M. (2003). It' about a time: research in challenges in digital archiving and long term preservation : Final report. The National Science Foundation and The Library of Congress, August 2003. http://www.si.umich.edu/digarch/NSF%200915031.pdf

Hess, C., & E. Ostrom. (2004, August 9-13, 2004). *Studying scholarly communication: can commons research and the IAD framework help illuminate complex dilemmas?* Paper presented at the 10th Biennial conference of the International Association for the Study of Commons Property, Oaxaca, Mexico. http://dlc.dlib.indiana.edu/archive/00001411/00/Hess_Studying_040512_Paper192.pdf

Hey, J. (2004). Targeting academic research with Southampton's institutional repository. *Ariadne*, 40. Available: http://www.ariadne.ac.uk/issue40/hey/

Hitchcock, S., T. Brody, J. M.N. Hey & L. Carr (2005). Preservation for Institutional Repositories: practical and invisible. *Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005)*, Edinburgh, November 21-23, 2005 http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/033.pdf

Hockx-Yu, H. (2006). Digital preservation in the context of institutional repositories. *Program: Electronic Library and Information Systems*, 40(3), 232-243.

Hoorn, E. & M. van der Graaf (2006). Copyright issues in Open Access research journals. *D-Lib Magazine, 12*(2), February 2006. Available: http://www.dlib.org/dlib/february06/vandergraaf/02vandergraaf.html

Hu, L. T. & P. M. Bentler (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.

Hurd, J. M. (1996). High Energy Physics. In *From print to electronic: the transformation of scientific communication* (pp. 9-34).

Hurd, J. M. (2000). The transformation of scientific communication: a model of 2020. *Journal of the American Society for Information Science, 51*(14), 1279-1283.

Jarvenpaa, S. L., & D. S. Staples. (2000). The use of collaborative electronic media for information sharing: an exploratory study of determinants. *Journal of Strategic Information Systems, 9*(2/3), 129-154.

Jones, R., T. Andrew, & J. MacColl. (2006). *The institutional repository*. Oxford, UK: Chandos Publishing.

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

Kennan, M. A. & C. Wilson (2006). Institutional Repositories: Review and an information systems perspective. *Library Management*, 27(4/5), 236-248.

Kim, H. (2001). The transition from paper to electronic journals: key factors that affect scholars' acceptance of electronic journals. *Serials Librarian, 41*(1), 31-64.

Kim, J. (2006). *Motivating and impeding factors affecting faculty contribution to institutional repositories.* Paper presented at the JCDL 2006 workshop: Digital curation and trusted repositories: seeking success, Chapel Hill, NC. http://sils.unc.edu/events/2006jcdl/digitalcuration/Kim-JCDLWorkshop2006.pdf

King, C. J., D. Harley, S. Earl-Novell, J. Arter, S. Lawrence, & I. Perciali. (2006). Scholarly communication: Academic values and sustainable models. Center for Studies in Higher Education, University of California, Berkeley, Berkeley, CA. http://cshe.berkeley.edu/publications/docs/scholarlycomm_report.pdf

Kling, R. (2004). The Internet and unrefereed scholarly publishing. *Annual Review of Information Science and Technology, 39*, 591-631.

Kling, R., & E. Callahan. (2003). Electronic journals, the Internet and scholarly communication. *Annual Review of Information Science and Technology, 37*, 127-177.

Kling, R., & R. Lamb. (1996). Analyzing visions of electronic publishing and digital libraries. In G. B. Newby & R. M. Peek (Eds.), *Scholarly Publishing: The Electronic Frontier*. Cambridge, MA: The MIT Press.

Kling, R., & G. McKim. (1999). Scholarly communication and the continuum of electronic publishing. *Journal of the American Society for Information Science, 50*(10), 890-906.

Kling, R., G. W. McKim, & A. King. (2003). A bit more IT: scholarly communication forums as Socio-Technical Interaction Networks. *Journal of American Society for Information Science and Technology, 54*(1), 47-67.

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

Kling, R., L. Spector, & G. McKim. (2002). Locally controlled scholarly publishing via the Internet: The Guild Model. *The Journal of Electronic Publishing, 8*. http://www.press.umich.edu/jep/08-01/kling.html

Knorr Cetina, K. (1999). *Epistemic cultures: how the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Karasti, H., K.S. Baker & E. Halkola. (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)*, 15(4): 321-358.

Krichel, T., & S. M. Warner. (2001). *Academic self-documentation: which way forward for computing, library and information science?* Paper presented at the 4th International Conference of Asian Digital Libraries, Bangalore, India. http://wotan.liu.edu/home/krichel/papers/mitaka.html

Kurtz, M. J., G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, E. Henneken, & S. S. Murray (2005). The effect of use and access on citations. *Information Processing & Management*, 41, 1395–1402.

Lancaster, F. W. (1995). Attitudes in academia toward feasibility and desirability of networked scholarly publishing. *Library Trends, 43*, 741-752.

Latour, B., & S. Woolgar. (1979). *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Lawal, I. (2002). Scholarly communication: the use and non-use of e-print archives for the dissemination of scientific information. *Issues in Science & Technology Librarianship, 36*. http://www.istl.org/02-fall/article3.html

Lawrence, S. (2001). Online or invisible? *Nature, 411*(6837), 521. http://citeseer.ist.psu.edu/online-nature01/

Liew, C. L., S. Foo, & K. R. Chennupati. (2000). A study of graduate student end-users' use and perception of electronic journals. *Online Information Review, 24*(4), 302-315.

Lynch, C. A. (1992). The transformation of scholarly communication and the role of the library in the age of networked information. *Serials Librarian, 23*(3/4), 5-20.

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL, 226*, 1-7.

Lynch, C. A., & J. K. Lippincott. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine, 11*(9). http://www.dlib.org/dlib/september05/lynch/09lynch.html

Lynch, C. (2007). The shape of the scientific article in the developing cyberinfrastructure. *CTWatch Quarterly*, 3(3), 5-10. http://www.ctwatch.org/quarterly/pdf/ctwatchquarterly-12.pdf

Manuel, K. (2001). The Place of e-prints in the publication patterns of physical scientists. *Science & Technology Libraries, 20*(1), 59-85.

Markey, K, Rieh, S.Y., St. Jean, B., Kim, J. & Yakel, E. (2007). Census of Institutional Repositories in the United States: MIRACLE project research findings: CLIR publication no.140, 167 p.

Mayernik, M. (2007) The prevalence of additional electronic features in pure e-journals. *The Journal of Electronic Publishing*, 10(3). http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;cc=jep;rgn=main;view=text;idno=3336451.0010.307

McVeigh, M. E. (2004). *Open Access journals in the ISI Citation Databases: Analysis of Impact Factors and citation patterns*. Thomson. http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf

Merton, R. K. (1973). The normative structure of science. In R. K. Merton & N. W. Storer (Eds.), *The sociology of science: theoretical and empirical investigations* (pp. 267-278). Chicago, IL: The University of Chicago Press.

Merton, R. K. (1988). The matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis, 79*(4), 606-623.

Meyer, E. T., & R. Kling. (2002). *Leveling the playing field, or expanding the bleachers? Socio-Technical Interaction Networks and arXiv.org*. Rob Kling Center for Social Informatics (RKCSI) Working Papers. https://scholarworks.iu.edu/dspace/handle/2022/149

Meyer, E. T. (2006). Socio-technical Interaction Networks: A discussion of the strengths, weaknesses, and future of Kling's STIN model. In Barleur J., Numinen M. I., Impagliazzo, J. (Eds). *IFIP International Federation for Information Processing Volume 223, Social Informatics: An Information Society for All? In Remembrance of Rob Kling* (pp.38-49). Boston: Springer. http://mypage.iu.edu/~etmeyer/files/HCC7_STIN_Meyer.pdf

Miles, M. B., & A. M. Huberman. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Miller, R. M. (2006). *Readers' attitudes to self-archiving in the U.K.* MSc Dissertation. School of Communication Arts, Napier University, Edinburgh, UK.

Moed, H. P. (2007). The effect of "Open Access" on citation impact: an Analysis of arXiv in Condensed Matter section. *Journal of American Society for Information Science and Technology, 58*(13), 2047-2054.

Molm, L. D. (1997). Social exchange and power. In *Coercive power in social exchange*. New York, NY: Cambridge University Press.

Nahapiet, J., & S. Ghoshal. (1998). Social capital, intellectual capital and organizational advantage. *Academy of management Review, 23*(2), 242-266.

Nicholas, D., P. Huntington, & H. R. Jamali. (2007). Open Access in context: A user study. *Journal of Documentation*, 63(6), 853-878.

Oakley, B., B. Pittman & T. Rudnick (2007). Tackling Copyright in the Digital Age: An Initiative of the University of Connecticut Libraries. UConn Library Published Works. http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1010&context=libr_pubs

Ober, J. (2005). *Postprint repository services: context and feasibility at the University of California*. Office of Scholarly Communication, University of California. http://osc.universityofcalifornia.edu/responses/materials/UC_postprintstudy_final.pdf

Odlyzko, A. M. (2002). The rapid evolution of scholarly communication. *Learned Publishing, 15*(1), 7-19.

Park, J., & J. Qin (2007). Exploring the Willingness of Scholars to Accept Open Access: A Grounded Theory Approach. *Journal of Scholarly Publishing*, 38(2), 55-84.

Pelizzari, E. (2003). Academic staff use, perception and expectations about open-access archives: a survey of Social Science Sector at Brescia University. http://eprints.rclis.org/archive/00000737/

Pelizzari, E. (2005). Harvesting for disseminating: open archives and the role of academic libraries. *The Acquisitions Librarian, 33/34*, 35-51.

Pinfield, S. (2001). How do physicists use an e-print archive? Implications for institutional e-print services. *D-Lib Magazine, 7*(12). http://www.dlib.org/dlib/december01/pinfield/12pinfield.html

Polydoratou, P. (2008). Use and linkage of Source and Output Repositories: Interviews with Chemistry researchers. *Aslib Proceedings* 60(1), 75-84.

Rieh, S. Y., Markey, K., Yakel, Y., St. Jean, B., & Kim, J. (2007). Perceived values and benefits of institutional repositories: A perspective of digital curation. An international symposium on Digital Curation (DigCCurr 2007), Chapel Hill, NC. April 18-20, 2007. http://www.ils.unc.edu/digccurr2007/papers/rieh_paper_6-2.pdf

RLG-OCLC (2002). Trusted Digital Repositories: Attributes and responsibilities. http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

Rowland, F., I. Bell, & C. Falconer. (1997). Human and economic factors affecting the acceptance of electronic journals by readers. *Canadian Journal of Communication, 22*(3/4), 61-75.

Rowlands, I., D. Nicholas, & P. Huntington. (2004). Scholarly communication in the digital environment: what do authors want? *Learned publishing*, 17, 261-273.

Rusch-Feja, D., & U. Siebeky. (1999). Evaluation of usage and acceptance of electronic journals. *D-Lib Magazine, 5*(10). http://www.dlib.org/dlib/october99/rusch-feja/10rusch-feja-full-report.html

Shearer, M. K. (2003). Institutional repositories: towards the identification of critical success factors. *Canadian Journal of Information and Library Science*, 27(3), 89-108.

Sismondo, S. (2004). Questioning functionalism in the sociology of science. In *Introduction to science and technology studies* (pp. 20-32). Oxford, UK: Blackwell Publishing Ltd.

Søndergaard, T. F., J. Andersen, & B. Hjørland. (2003). Documents and the communication of scientific and scholarly information: revising and updating the UNISIST model. *Journal of Documentation, 59*(3), 278-320.

Speier, C., J. Palmer, D. Wren, & S. Hahn. (1999). Faculty perceptions of electronic journals as scholarly communication: a question of prestige and legitimacy. *Journal of the American Society for Information Science, 50*(6), 537-543.

Stewart, L. (1996). User Acceptance of Electronic Journals: Interviews with Chemists at Cornell University. *College & Research Libraries, 57*, 339-349.

Suber, P. (2002). Open access to the scientific journal literature. *Journal of Biology, 1*(1). http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=117246&blobtype=pdf

Swan, A., & S. Brown. (2004). *ISC/OSI Journal author survey report*: JISC, HEFCE. http://eprints.ecs.soton.ac.uk/11002/01/JISCOAreport1.pdf

Swan, A., & S. Brown. (2005). *Open access self-archiving: an author study*. Key Perspective Ltd. http://eprints.ecs.soton.ac.uk/11006/01/jiscsum.pdf

Sweeney, A. E. (2000). Tenure and promotion: should you publish in electronic journals? *The Journal of Electronic Publishing, 6*. http://www.press.umich.edu/jep/06-02/sweeney.html

Taubes, G. (1996). Electric Preprints Point the way to 'Author Empowerment'. *Science, 271*(5250), 767-768.

Tschider, C. (2006). Investigating the 'public' in the Public Library of Science: Gifting economics in the Internet community. *First Monday*, 11(6). http://firstmonday.org/issues/issue11_6/tschider/index.html

The University of California Office of Scholarly Communication & The California Digital Library e-Scholarship Program (2007). Faculty attitudes and behaviors regarding scholarly communication: Survey findings from the University of California. http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf

Van de Sompel, H., & C. Lagoze. (2000). The Santa Fe convention of the open archives initiative. *D-Lib Magazine, 6*(2). http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

Van House, N. A. (2002). Digital libraries and practice of trust: networked biodiversity information. *Social Epistemology,* 16(1), 99-114.

Van House, N. A. (2003). Digital libraries and collaborative knowledge construction. In A. P. Bishop, N. A. V. House & B. P.Buttenfield (Eds.), *Digital library use: social practice in design and evaluation* (pp. 271-295). Cambridge, MA: MIT Press.

Van House, N. A., A. P. Bishop, & B. P.Buttenfield (2003). Introduction: digital libraries as sociotechnical systems. In A. P. Bishop, N. A. V. House & B. P.Buttenfield (Eds.), *Digital library use: social practice in design and evaluation* (pp. 1-24). Cambridge, MA: MIT Press

Van Raan, A. F. J. (1997). The future of the quality assurance system: its impact on the social and professional recognition of scientists in the era of electronic publishing. *Journal of Information Science, 23*(6), 445-450.

Ware, M. (2004). Institutional repository and scholarly publishing. *Learned publishing, 17*, 115-124.

Watson, S. (2007). Authors' attitudes to, and awareness and use of, a university Institutional Repository." *Serials: The Journal for the Serials Community* 20(3), 225-230.

Wojciechowska, A. (2007). Analysis of the use of Open Archives in the fields of Mathematics and Computer Science." *OCLC Systems & Services*, 23(1), 54-69.

Warlick, S. E. & K.T.L. Vaughan (2007). Factors influencing publication choice: Why faculty choose open access. *Biomedical Digital Libraries*, 4(1). http://www.bio-diglib.com/content/4/1/1

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

Wasko, M. M., & S. Faraj. (2005). Why should I share? Examining knowledge contribution in networks of practice. *MIS Quarterly*, *29*(1), 35-57.

Willinsky, J. (2006). *The access principle: the case for open access to research and scholarship*. Cambridge, MA: MIT Press.

Yakel, E., Rieh, S.Y., St. Jean, B., Markey, K., and Kim, J. (2008). Institutional repositories and the institutional repository. *American Archivist*, 71(2).

Young, J. R. (2002). 'Superarchive' could hold all scholarly output. *The Chronicle of Higher Education*, 48(43). http://chronicle.com/free/v48/i43/43a02901.htm

Xia, J. (2007). Factors to Assess Self-Archiving in Institutional Repositories. *Serials Review*, 33(2), 73-80.

Ziman, J. (1984). Authority. In *An introduction to science study: the philosophical and social aspects of science and technology* (pp. 70-80). New York, NY: Cambridge University Press.

Zimmerman, A. (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7(1-2), 5-16.