Socio-Technical Factors of Scholarly Communication in an Institutional Repository Context

Nathan Hall
Center for Digital Research and Scholarship, University Libraries
Virginia Tech, USA
nfhall@vt.edu

ABSTRACT
This exploratory study uses a phenomenological approach to examine faculty attitudes in an institutional repository context as well as the socio-technical factors that affect scholarly communication and data sharing practices. The author conducts interviews and observations with Environmental Studies faculty at two large public research universities. This article reports preliminary results, though data transcription and analysis are ongoing. The author plans to analyze interview transcriptions to find dominant themes. The themes will inform recommendations to improve institutional repository services for the stakeholders.

Categories and Subject Descriptors
J.4 [**Social and Behavioral Sciences**]: *Economics, psychology, sociology*

General Terms
Human Factors

Keywords
Data sharing, Scholarly Communication, Socio-Technical Factors, Research Environments, Phenomenology, Institutional Repositories

1. INTRODUCTION

1.1 Background
New information technologies have created new expectations in scholarly communication, including the accessibility of research data [1][2]. Policies and culture have shifted both locally and internationally in response to new capabilities. Since the National Institutes of Health released the Final NIH Statement on Data Sharing in 2003 [3], several other funding agencies have followed suit and provided guidelines for providing open access for research data and peer-reviewed articles resulting from sponsored research. Proposed legislation such as the Fair Access to Science and Technology Research Act, and federal mandates such as the Policy Memorandum for Increasing Access to the Results of Federally Funded Scientific Research from the White House Office of Science and Technology Policy [4] have increased expectations for faculty to share the results of federally funded research. This is an opportunity for institutional repositories and other digital curation programs to provide services to help researchers meet those demands, yet there are still problems with low deposit rates for institutional repositories.

Lynch [5] defines institutional repositories 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" and states that they could or should contain research and teaching materials as well as experimental and observational data that support the scholarly activities of researchers at the institution.

1.2 Research Problem
Teaching and research faculty attitudes towards institutional repository services are poorly understood. Repository developers, managers, and administrators have been largely disconnected from their faculty stakeholders needs in many universities [6]. As a result, institutional repositories and the Open Access movement have not been as successful on a national scale in the United States as early advocates hoped [7][8].

Some existing studies explore faculty deposit patterns [9] and attitudes [10], including a study that examines the very notion of "authorship" and responsibility of research data [11] but the literature review did not find any studies that specifically examine socio-technical factors of faculty attitudes regarding scholarly communication and data sharing in an institutional repository context.

A shifting policy landscape will impact the scholarly communication and data sharing practices of faculty, particularly for federally funded research, but that does not mean that academic libraries can be complacent about addressing stakeholder concerns in the development and deployment of institutional repositories. Academic libraries have new opportunities to become partners in the campus research environment, but a better understanding of faculty attitudes and the socio-technical factors that affect scholarly communication and data sharing practices will help academic libraries provide better services.

1.3 Research Questions
1. How do teaching and research faculty want to disseminate their research?
2. What attitudes do teaching and research faculty have towards data sharing?
3. What socio-technical factors promote or inhibit faculty adoption of institutional repositories?

1.4 Research Goals
The goals of this research are to:

- Determine the socio-technical factors that affect faculty attitudes towards institutional repository services.
- Produce findings that will inform the future development of institutional repository services, particularly for the stakeholders in this project.
- Produce findings that will inform academic library strategies for partnering with teaching and research faculty.
- Develop areas for further research to produce generalizable findings.

2. STATE OF THE FIELD

To provide a framework for understanding the attitudes of researchers in an institutional repository context, this study relies on the body of work in several different areas of research, including information behavior, scholarly communication, data management, and innovation diffusion theory.

Information behavior studies cover a wide range of activity [12][13]. One study found discipline, time spent in an organization, task role, and cognitive style each has an impact on information behavior [14]. Another author used ethnography to study a digital library community, an approach which emphasized the importance of "ongoing investigation of emergent phenomena" to track the development of a digital library and its stakeholders [15].

Research on digital curation of research data, particularly the IMLS funded Data Curation Profiles project at the Distributed Data Curation Center at Purdue University Libraries is useful for the study of how faculty manage research data. This project is still underway but the process for designing the tool is documented [16]. Additionally, the completed Data Curation Profiles (http://datacurationprofiles.org/completed) provide a data set for potential analysis, though no studies have been published yet. Further infrastructure and training for librarians is necessary for more robust data curation services in academic libraries [17].

A survey conducted at California Polytechnic Institute at San Luis Obispo found that "while the majority of researchers believe that colleagues should share their data, only a minority of respondents actually share their own data with individuals who did not help in gathering the data" [18]. The question of why researchers do not share their data was not within the scope of the study.

One qualitative study by Foster and Gibbons [6] at the University of Rochester used work-place observations and interviews with faculty in various disciplines to determine how an institutional repository could support existing work practices. The co-authors developed marketing strategies to better communicate benefits of the institutional repository to faculty researchers. They also developed customized researcher pages to promote the works of individual faculty members rather than the output of the institution. Foster and Gibbons demonstrate the value of qualitative methods to better understand faculty and for generating findings to improve repository services. The study however does not examine how socio-technical factors affect faculty use of institutional repositories.

User acceptance of technology and innovation diffusion theory provide a theoretical background for studying socio-technical factors in institutional repository contexts [19]. Dillon and Morris cite three factor groups that influence technology acceptance, including user psychology, information technology design process, and user perception of quality.

The study of socio-technical factors was useful in a study of scholarly communication forums [20]. The authors found that when research and development only examines technical factors such as information processing features and only examines users as individual actors, then the cultural work context and the information ecology is ignored. Technological development does not serve sustainability or user-acceptance unless developers and administrators also include these additional social and behavioral factors when analyzing user requirements. The study found that a socio-technical interaction network-based model forces social analysis into the entire process of development and management, instead of just in planning and evaluation.

3. RESEARCH DESIGN

3.1 Strategy of Inquiry

This study uses a phenomenological approach in the collection and analysis of qualitative data from semi-structured interviews to determine socio-technical factors that inhibit faculty adoption of institutional repositories at the Institute of Applied Science at the University of North Texas, and at a few selected colleges that study environmental issues at Virginia Tech.

Phenomenology is a research strategy to understand "human experiences about a phenomenon as described by participants" [21]. Phenomenological research involves prolonged individual contact with a small number of participants to identify common trends. Phenomenology is often characterized by in depth qualitative data from a very small sample. Its value comes from its emphasis on personal perspective and subjective experience. It can identify deep issues that would be easily overlooked by other methods, such as surveys or log analysis. The limitation from the small sample size in phenomenology is that it is difficult to produce generalizable findings, but it conversely provides a route to understanding "the meaning attributed by persons to the activities in which they engage, in order to understand their behavior" [22].

Phenomenology is an appropriate strategy for this research for a variety of reasons. It is flexible compared to other approaches. Flexibility is a critical feature since there is no established general tool for understanding user communities in an institutional repository context. Institutional repositories at the two locations in this study are still in the early stages of development. The two populations in question are not well understood in terms of their information searching behavior and a study of these groups would be useful to the designers of these technologies and services. Because of these factors, phenomenology is a practical exploratory approach for gathering data about how these researchers seek information, and how access to and management of these information resources could be improved.

Like with other qualitative research methods, the use of a phenomenological approach affects the questions asked because depending on a participant's answer to a question in the interview, the researcher asks probing follow-up questions that will not have necessarily been planned.

A phenomenological approach affects data collection in a number of ways. Data generated from audio-recorded semi-structured interviews are descriptive rather than numeric. Data collection occurs in a natural setting rather in a controlled research environment and, in this study, timetables were subject to the schedules of the participants.

The use of a phenomenological approach affects data analysis because it involves data collected from human sources. Data analysis is subject to nuanced interpretation as a researcher attempts to generalize and code observed behaviors and recorded quotations.

Finally, the phenomenological strategy affects the final narrative by omitting some details that are irrelevant to the research problem. Not all collected data is useful to the stakeholders. A researcher must determine what data trends are relevant based on how clearly they represent participant attitudes and how clearly the relate to the research problem.

3.2 Researcher's Role

Because this research deals with human participants, it is necessary to obtain permission to conduct the study from the

respective Institutional Review Boards at Virginia Tech and UNT. I submitted my list of interview questions to them, and they approved the project.

I am connected to the University of North Texas participants in this study as a former colleague and as a currently enrolled doctoral candidate. I am connected to the Virginia Tech participants as a current faculty member of that university. Having that connection to the communities facilitated the process of recruiting participants. I am connected to UNT Libraries as a former librarian. I am connected to the VTechWorks repository as the coordinator of that project and I work closely with the software developers.

Due to my connection to the two digital library projects, risk of bias on my part is significant. I manage the level of bias by avoiding the sources listed by Schensul et al. [23]. These include asking leading questions, failing to follow up on topics introduced by the participant, interrupting or redirecting the participant, failing to recognize participant reactions to my appearance, using nonverbal cues that encourage the participant to give certain answers, or voicing my opinions on any issues.

I recruited participants by two methods. I sent an email to their work email address, requesting appointments to discuss their research activities. The email stated that this study is a doctoral dissertation project with the objective of improving the information resources available to environmental policy researchers. I also used my relationships to the communities to identify potential participants through snowball sampling. The appointments included a semi-structured interview over the phone, or VOIP, or in person, depending on their location and their relative convenience and comfort.

The participants will benefit from this study by providing input to improve a research tool that serves their research needs. The results will be reported in the final draft of the dissertation, which will be freely available through the websites of the UNT Libraries and Virginia Tech University Libraries. All participants will be informed when the study is available.

Identifying characteristics of the participants will be masked in all sections of the report. Participants were assigned numbers to protect their identities. I also recorded certain profile information, such as whether the participant was tenured or not.

The participants will benefit from this study by providing input to improve a research tool that serves their research needs. The results will be reported in the final draft of the dissertation, which will be freely available through the websites of the UNT Libraries and Virginia Tech University Libraries. All participants will be informed when the study is available.

3.3 Data Collection Procedures

3.3.1 Bounding the Study

While interviewing researchers in other disciplines, or including graduate students, would produce useful findings, it is convenient to limit the research scope in this exploratory phase. Environmental Studies is a suitable field due to its interdisciplinary nature. Environmental Studies researchers are diverse in methodology, philosophy, tools, culture, and practice. The discipline includes scientists, humanists, and social scientists. Robust Environmental Studies programs exist at both sites, which means there is a convenient sample.

3.3.2 Setting

I conducted the study at two locations. One location is the Denton campus of the University of North Texas. This is a large state university with between 35,000 and 40,000 students seeking degrees. The campus is the site of the Institute of Applied Science, an interdisciplinary research unit focused on human interaction with the environment, and solutions to environmental problems. The other location is the Blacksburg campus of Virginia Tech. Virginia Tech is a large state university with nearly 30,000 students in degree-seeking programs. Several units at Virginia Tech have faculty in relevant research areas. These include the College of Natural Resources and Environment, the College of Liberal Arts and Human Sciences, the College of Agriculture and Life Sciences, and the Department of Civil and Environmental Engineering. Both universities have teaching and research faculty and graduate students focused on education, outreach, and research directed at environmental problems, sustainability, conservation, and human interaction with the environment and its natural resources.

3.3.3 Actors

I interviewed faculty at the UNT Institute of Applied Sciences. The faculty members of the Institute of Applied Sciences are a target audience of potential users of the Scholarly Works Collection in the UNT Digital Library.

I interviewed faculty at Virginia Tech in relevant areas of research and instruction. The faculty members are a subgroup of potential users of the VTechWorks Institutional Repository.

3.3.4 Events

During the interviews, I asked the participants in the study questions about their research and instructional activities, their use of digital libraries and repositories, and their perceived barriers to access and adoption.

3.3.5 Process

The study gathered data through semi-structured interviews. I asked questions of the participants, preferably face-to-face, though in some cases due to the travel schedules of faculty members some interviews may take place using VOiP software or a telephone. I recorded data with a digital audio recorder for interviews, and I transcribed interviews. I also took descriptive and reflective notes during the interviews.

3.4 Data Analysis and Interpretation

Data analysis will occur through an iterative process of identifying and labeling common themes in the interviews and observations. The first step for analyzing the interview data is transcribing recorded interviews. Then I will use the qualitative research software NVivo to help me annotate transcriptions with terms representing concepts that I discover. Using word search features, I will add concept terms and apply them throughout the transcribed interviews wherever I believe those concepts occur. I will organize occurrences of the concept terms into themes.

The dominant themes that emerge from the interviews will be the main findings of this study.

3.5 Validation

Validation of findings in this study will occur through several strategies—triangulation, member-checking, and clarification of bias.

Triangulation occurs by collecting data from two sites in order to justify the themes defined in data analysis and interpretation. I will engage in member-checking by sharing the themes with the interview participants before I publish the final findings to determine whether they feel the themes accurately represent what they expressed in the interviews and felt during the observations. I will include their views of the findings in the final report. I will clarify the bias that I bring to the study by explaining my relationship to the participants and to the institutions that are the sites of the research.

4. PRELIMINARY FINDINGS

Informal analysis has only covered the data sharing aspects of the study thus far. Early analysis indicates a number of overlapping, and at times conflicting views about providing free online access to research data.

Nearly all participants see value in sharing research data. The faculty members who saw little value in sharing their data believed that their data was too narrow to be useful in a different study.

Other reasons for hesitating to share research data included concern about IRB compliance and the protection of research subjects. Another common reason some participants did not want to share their data was the amount of work required to organize the data to make it useful to someone else.

While many researchers felt that sharing data was necessary for transparency of the scientific process, some were concerned that their data could be taken out of context and politicized. These participants were all senior faculty.

Faculty who already have tenure were more likely to speak favorably about data sharing to improve the transparency of the scientific process and some indicated they would appreciate having any impact at all (on environmental policy, for example). Many junior faculty members expressed concern about their ability to publish the findings on their own data before someone else could.

Interview transcription is ongoing and has not undergone any formal clustering or post-collection validation. The conclusions drawn in this preliminary reading will be subjected to member-checking and may guide future research design in order to gather data from larger sample and produce generalizable findings.

5. EXPECTED CONTRIBUTIONS

This study confines itself to interviews and observations of environmental studies faculty at two large, PhD granting, public research universities. The sample and the setting reduce the generalizability of the findings. Furthermore, as a qualitative study the data could produce alternative findings and be subject to alternative interpretations.

The preliminary findings support and build on the conclusions of related studies [24][25], though most of the participants in this study seemed to be more aware of and expressed attitudes more in favor of open access publishing than the participants in the two cited studies. Whether this difference reflects changing attitudes towards open access in the last several years, or if it is a reflection of the sample is not determined.

The major contributions of the research will fall under two categories—practical and theoretical.

The practical contributions will include recommendations for incentive mechanisms to foster faculty acceptance and use of library services to provide open access dissemination of research findings through data sharing and repository deposit. Such recommendations may include more specific guidelines from funding agencies so that principle investigators better understand the expectations. The recommendations will also include improvements to institutional repository interfaces and infrastructure so that institutional repositories can automatically discover and harvest faculty research. Institutions with Open Access mandates in place could use discovery and harvest tools to automatically populate repositories so that the onus of depositing is not on the researcher but on the institution.

The research will be useful to institutions developing institutional repository services by helping help developers and administrators understand the priorities of faculty partners and stakeholders. The research will also serve liaison librarians who work closely with environmental studies faculty. The findings will help both commercial publishers and advocates for open access publishing understand a key interdisciplinary constituency.

The theoretical contributions of this study may add to discussions on Theory of Reasoned Action and social-technical systems theory.

Future research will build on this study and include the use of a different research paradigm, strategy, and methodology in order to produce more generalizable findings.

6. REFERENCES

[1] Association of Research Libraries, "Scholarly Communication." Association of Research Libraries.

[2] Association of Research Libraries, "To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering." Association of Research Libraries, Sep-2006.

[3] National Institutes of Health, "Final NIH Statement on Sharing Research Data, NOT-OD-03-032," 26-Feb-2003. [Online]. Available: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html. [Accessed: 15-Apr-2013].

[4] J. P. Holdren, "Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research." Executive Office of the President, Office of Science and Technology Policy, 22-Feb-2013.

[5] C. A. Lynch, "Institutional repositories: Essential infrastructure for scholarship in the digital age.," *Arl Bimon. Rep.*, no. 226, Feb. 2003.

[6] N. F. Foster and S. Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories," *D-Lib Mag.*, vol. 11, no. 1, Jan. 2005.

[7] D. Salo, "Innkeeper at the Roach Motel," *Libr. Trends*, vol. 57, no. 2, pp. 98–123, 2008.

[8] C. S. McDowell, "Evaluating Institutional Repository Deployment in American Academe Since Early 2005 Repositories by the Numbers, Part 2," *D-Lib Mag.*, vol. 13, no. 9/10, Oct. 2007.

[9] C. Thomas and R. H. McDonald, "Measuring and Comparing Participation Patterns in Digital Repositories: Repositories by the Numbers, Part 1," *D-Lib Mag.*, vol. 13, no. 9/10, Oct. 2007.

[10] P. M. Davis and M. J. L. Conolly, "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace," *D-Lib Mag.*, vol. 13, no. 3/4, Apr. 2007.

[11] J. C. Wallis and C. L. Borgman, “Who is Responsible for Data? An Exploratory Study or Data Authorship, Ownership, and Responsibility,” *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 48, no. 1, pp. 1–10, Oct. 2011.

[12] Karen E. Fisher and Heidi Julien, “Information behavior,” *Annu. Rev. Inf. Sci. Technol.*, vol. 43, no. 1, pp. 1–73, 2009.

[13] D. O. Case, “Information Behavior,” *Annu. Rev. Inf. Sci. Technol.*, vol. 40, no. 1, pp. 293–327, 2006.

[14] J. Palmer, “Scientists and information: II. personal factors in information behavior,” *J. Doc.*, vol. 47, no. 3, pp. 254–275, 1991.

[15] M. Khoo, “Ethnography, Evaluation, and Design as Integrated Strategies: A Case Study from WES,” in *Research and Advanced Technoogy for Digital Libraries*, London, U.K.: Springer-Verlag, 2001, pp. 263–274.

[16] M. Witt, J. Carlson, D. S. Brandt, and M. H. Cragin, “Constructing Data Curation Profiles,” *Int. J. Digit. Curation*, vol. 4, no. 3, pp. 93–103, 2009.

[17] J. Carlson, “Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data,” *Ref. Serv. Rev.*, vol. 40, no. 1, pp. 7–23, 2012.

[18] J. M. Scaramozzino, M. L. Ramírez, and K. J. McGaughey, “A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University,” *Coll. Res. Libr.*, vol. 73, no. 4, pp. 349–365, Jul. 2012.

[19] A. Dillon and M. Morris, “User Acceptance of Information Technology:Theories and Models,” in *Annual Review of Information Science and Technology*, vol. 31, M. Williams, Ed. Medford, NJ: Information Today, 1996, pp. 3–32.

[20] R. Kling, G. McKim, and A. King, “A bit more to it: Scholarly Communication Forums as Socio-Technical Interaction Networks,” *J. Am. Soc. Inf. Sci.*, vol. 54, no. 1, pp. 47–67, Jan. 2003.

[21] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd ed. Thousand Oaks, California: Sage Publications, 2003.

[22] T. D. Wilson, “Alfred Schutz, phenomenology and research methodology for information behaviour research,” *Isic4 - Fourth Int. Conf. Inf. Seek. Context*, Sep. 2002.

[23] S. L. Schensul, *Essential ethnographic methods: observations, interviews, and questionnaires*, vol. 2. Rowman Altamira, 1999.

[24] D. Harley, S. Earl-Novell, J. Arter, S. Lawrence, and C. J. King, “The Influence of Academic Values on Scholary Publication and Communication Practices.,” *J. Electron. Publ.*, vol. 10, no. 2, Spring 2007.

[25] The University of California Office of Scholarly Communication, The California Digital Library eScholarship Program, and Greenhouse Associates, “Faculty Attitudes and Behaviors Regarding Scholarly Communication: Survey Findings from the University of California,” Aug. 2007.