

# Document Practice as Insight to Digital Infrastructures of Distributed, Collaborative Social Scientists

Sarika Sharma<sup>1</sup>, Jaime Snyder<sup>1</sup>, Carsten Østerlund<sup>1</sup>, Matt Willis<sup>1</sup>, Steven Sawyer<sup>1</sup>, Michelle Brown<sup>1</sup> and Dorotea Szkolar<sup>1</sup>

<sup>1</sup> Syracuse University

## Abstract

The objective of this study is to understand scholarly research practice in virtual, distributed collaborations by focusing on the flow of documents among the participants and to advance design guidance for supporting improved document practice across distributed collaborative platforms. To do so, we develop a theoretical framework on document practice highlighting the sociotechnical role of documents in digital infrastructure. This mixed-methods study will first conduct semi-structured interviews to understand document practices. The second phase of the study will collect trace data of documents as a way to understand how they change over time. In this poster, we report on the analysis of twelve interviews from social scientists working in virtual collaborations. Initial findings show that social scientists organize their documents and scholarly work on emergent digital infrastructures. Although not ideal, emergent digital infrastructures provide stability for collaboration across time and space.

**Keywords:** virtual organizing, social science, cyberinfrastructure

**Citation:** Sharma, S., Snyder, J., Østerlund, C., Willis, M., Sawyer, S., Brown, M., & Szkolar, D. (2014). Document Practice as Insight to Digital Infrastructures of Distributed, Collaborative Social Scientists. In *iConference 2014 Proceedings* (p. 1021–1024). doi:10.9776/14355

**Copyright:** Copyright is held by the authors.

**Acknowledgements:** This material is based upon work supported by the National Science Foundation under Grant No. OCI-12219445. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This material is based upon work supported by the National Science Foundation under Grant No. OCI-12219445; the Institute for Museum and Library Science, and the iSchool at Syracuse University

**Contact:** skshar01@syr.edu, jasnyd01@syr.edu, coesterlund@syr.edu, mawillis@syr.edu, ssawyer@syr.edu, mlbrow11@syr.edu, doroteaszskolar@syr.edu

## 1 Introduction

Scholarly academic research is becoming more distributed and collaborative as information communication technologies make it possible to collaborate, coordinate and organize scholarly work (Cummings and Kiesler, 2005 and Palmer, Tefteau, and Pirmann, 2009). Collaborative science is conceptualized by the size and shape of its collaborative members. There is a relationship between the size of the collaboration and the resources necessary to coordinate and organize work across collaborations (Cragin, Palmer, Carlson and Witt, 2010). Collaborative science may be small-scale, medium-scale, or large-scale in size. Small-scale collaborations are understood as small science (Cragin, et al., 2010). These collaborations are comprised of several principal investigators who work together on research grants (Cragin et al., 2010). On the other hand, large-collaborations are emphasized as big science or collaborations with hundreds of researchers and dispersed resources. Using this framework to understand digital infrastructures highlights the majority of work on cyberinfrastructure to support large-scale collaborations.

The literature in cyberinfrastructure (CI) studies conceptualizes the common components necessary to build and maintain a research infrastructure. Ribes and Lee have summarized seven main facets to understand cyberinfrastructure in large-scale academic research in the natural sciences and computer science. These seven facets serve as important sources to look when understanding digital infrastructures.

The first facet of cyberinfrastructure is to observe it as an infrastructure or a structured platform that is maintained over time (Ribes and Lee, 2010). A way to analyze infrastructure is to follow people and technology and observe areas of strain and areas of success (Ribes and Lee, 2010). This requires looking at the technical work and the human work to understand how the infrastructure is being sustained, maintained and how it may scale-up (Ribes and Less, 2010). These facets of CI may be useful to understand the digital infrastructures of small-scale distributed, collaborative research, specifically, the social use of documents and the technical use of documents.

The purpose of this research is to follow distributed collaborative social scientists to observe their digital infrastructures by looking at their documents. Documents provide insight into academic research work because they serve as important artifacts throughout the research phase. Previous work shows the important role of documents in academic research. Brady and Berman (2005) describe social science cyberinfrastructure as “multidimensional networks that include individuals, data sets, documents, analytic tools, and concepts.” Palmer et al. (2009) describes documents as key resources to scholarly academic work. To date, very few studies have looked at document practice or documents as a stable artifact to help understand infrastructure. In this study, document practice is defined as the action or performance of or with a document pertaining to using, sharing, combining, and storing. Traditionally, documents in research work are defined as manuals, papers, journal articles etc. (Palmer, 2009). Both document practice and documents are understood from the perspective of the participant thus informing ground-up theory formation. This research study will try to understand the sociotechnical nature of documents as a way to elucidate digital infrastructures in academic research.

## 2 Methods

To date, we have conducted semi-structured interviews with twelve social scientists part of distributed collaborative research teams. Our interview protocol also included a structured checklist that captured details about digital information and communication tools used by researchers in doing their work online. Our sample was purposive consisting of twelve social scientists from the information studies tradition. Our sample criteria included researchers who were working on currently funded distributed collaborative research projects. Our rationale for selecting information scientists included two main factors. First, we believed that if any social scientists were going to be using digital infrastructures for collaboration, these technology-oriented scholars would represent “power users.” Second, because many of the information scientists we spoke with conduct their own research in the area of virtual collaboration and information technologies, we felt they would have a particularly generative ability to reflect on and articulate their practices. Interviews were transcribed, inductively coded and analyzed for themes pertaining to document practices and arrangements of digital infrastructure (Charmaz, 2006). Results from coding were compared to checklist data regarding the types of tools and licensing that are commonly used for distributed collaboration.

## 3 Findings

The interviews provided some interesting insights on participants documenting practices and the technical platforms they use to conduct scholarly work. Distributed collaborative research in social science is clearly driven by documents. Across all the interviews, documents were the primary means to conduct research work. All twelve interviewees mentioned one or more of the following documents as they described their collaborative work: Word documents, PowerPoint slides, PDFs, email, and wikis. These were also the most reported documents in collaborations. When asked about their documenting practices pertaining to these documents, almost all participants described their collaborative work around a combination of editing, sharing, using, creating, versioning and/or drafting documents. Genres of these documents spanned from paper drafts, field notes, and annotations, to project related files.

Second, based on the checklists, when it comes to storing, accessing, exchanging documents, research teams relied on commercial technology to conduct most of their document work. Respondents described their documenting practices as being associated with certain technologies. For instance, participants said that Email is the primary tool for exchanging, sharing, and interacting for almost all project work. Email was the most common online tool across all of our respondents for exchanging documents. Respondents reported using software and hardware such as Microsoft Office applications, browsers such as Firefox or Safari and some common open source programs such as “R,” running on Microsoft or Apple computers. To store documents, several participants used DropBox and/or Google Drive. Collaboration on documents primarily took place via Word document exchange on email or Google Drive through Google documents. Commoditized technologies such as Google Drive, and Dropbox became spaces for storing documents, managing documents, and collaborating with documents. Skype was primarily used as means to communicate with geographically dislocated members. Several participants noted the dependency of project phase to certain documenting practices and technologies associated with those practices. In this case, programs were added and dropped throughout the duration of projects on a needs base. Some documenting practices were not associated with certain technologies. For example, when describing the versioning of documents during writing phases some displayed interesting versioning processes. Two respondents described token passing, some also described titling documents, while many relied on email as a passing back and forth tool. One respondent talked about the lack of interest in the organization to keep up with wikis and technologies used to organize documents.

When looking at the programs and technologies respondents used, they varied greatly across collaborations. Participants described various conglomerations of technology combinations and no two groups had similar digital infrastructures. Some participants described how their arrangements were compiled while some described their ideal arrangements. For example, one participant concerned with group dynamics chose to use wikis because they are “democratic” in flavor and they serve as spaces where various documenting practices can take place from storing data to writing/compiling annotations. Others wished they had spent more time understanding technologies that could help their document practice. Others displayed strong dislike for Google documents. Several participants mentioned that using Google drive is not a preferred way of collaborating and organizing online but it is currently the best option to situate and coordinate writing research papers with many researchers. All of participants described their lack of project management or governing bodies to help with technology arrangements to provide seamless foundation on which work could be done. One respondent described the way they choose their suite of technology arrangements. In this case, the participant looked for programs that required very little learning so as to not take time away from project goals.

## 4 Discussion

The document practice lense provides insight on the types of documents and the ways in which documents are used in collaborations. Our respondents described their collaboration on documents and ways in which they interact with group members. From this insight, documents provide a base for discussion on the ways researchers work online. Document practice identifies some of the basic actions that are used when working with documents online. Insights from our interview respondents display some of the areas of where the social meet with the technical aspects of documents highlighting digital infrastructures.

The digital infrastructures that we observed from respondents are very different from the characteristics of CI noted in the literature. The distributed collaborative activities of our respondent’s occur on cobbled-up commercial technologies. Their infrastructures are comprised of commercial software, hardware, and Internet based services to store documents, interact with dislocated colleagues, and write paper drafts. Modularity of the technologies gave collaborations stability and flexibility. Social Scientists in our study did not discuss the use of computational power to analyze data or reflect on robust data sharing

mechanisms to conduct work. What they did say was the need for platforms where work can be shared and stored across researchers seamlessly. This entails being able to coordinate collaboration on documents especially for paper writing papers. More digital infrastructures are compiled to take documenting practices into consideration that may reflect the importance of the types of work that is being conducted in distributed collaborative social science.

Although digital infrastructures do not reflect the model of CI proposed by Atkins, 2003, they do provide similar functions that allow collaborations to survive and thrive for the duration of the project. They provide stable yet fluid and evolving platforms in which distributed work successfully takes place. What is surprising to notice is that these compiled information and communications technology (ICT) do not have governance structures or support systems to maintain them; rather they are reflections on the skills of the person who compiled them. These findings suggest that social scientists are inventively appropriating ICT use into their scientific practice.

## 5 Conclusion

In the next phase of the study, we wonder what technical arrangements provide the most stability for document practices? Are there further insights we can get from looking at documents in their digital infrastructures? How do technology arrangements look over time? These are some of the questions we look to answer in the next phase of the study. In terms of methodology, the next phase of our study will be to develop a trace data protocol in order to collect data on how documents flow in research collaborations. Our work hopes to forward theoretical understanding of the document practice to further understand its application to forming digital infrastructures in scholarly research.

## 6 References

- Atkins, D.E., Droegemeier, K.K., Feldman, S.I., Garcia-molina, H., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P. and Wright, M.H. Revolutionizing Science and Engineering Through Cyberinfrastructure : Report of the National Science Foundation Blue-Ribbon Advisory Panel on cyberinfrastructure. *Science*, 81 (8). 1562-1567.
- Berman, F. D., & Brady, H. E. (2005). Final report: NSF SBE-CISE workshop on cyberinfrastructure and the social sciences. Arlington, VA: National Science Foundation.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Pine Forge Press.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038.
- Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, 35(5), 703-722.
- Kaziunas, E., Sawyer, S., & Østerlund, C. (2011, February). Social scientists, documents and cyberinfrastructure: the cobbler's children or the missing masses?. In *Proceedings of the 2011 iConference* (pp. 166-173). ACM.
- Østerlund, C. S. (2008). Documents in place: demarcating places for collaboration in healthcare settings. *Computer Supported Cooperative Work (CSCW)*, 17(2-3), 195-225.
- Palmer, C. L., Tefteau, L. C., & Pirmann, C. M. (2009). Scholarly information practices in the online environment. *Report commissioned by OCLC Research. Published online at: [www.oclc.org/programs/publications/reports/2009-02.pdf](http://www.oclc.org/programs/publications/reports/2009-02.pdf)*.
- Ribes, D., & Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-Research: Current themes and future trajectories. *Computer Supported Cooperative Work (CSCW)*, 19(3-4), 231-244.