Documents and Distributed Scientific Collaboration

Matt Willis
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
mwillis@syr.edu

Michelle Brown
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
mlbrow11@syr.edu

Sarika Sharma
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
skshar01@syr.edu

Carsten Østerlund
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
costerlu@syr.edu

Jaime Snyder
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
jasnyd01@syr.edu

Steve Sawyer
School of Information Studies
Syracuse University
900 S Crouse Ave.
Syracuse, NY 13210, USA
ssawyer@syr.edu

CSCW'14 Companion, February 15–19, 2014, Baltimore, MD, USA.
ACM 978-1-4503-2541-7/14/02.
http://dx.doi.org/10.1145/2556420.2556491

Abstract
We ask the question: *What document infrastructures do scientists build to support their virtual organizing and documenting practices?* Cyberinfrastructure (CI) is seen by many as playing a critical role in the future of social, behavioral, and economic sciences (SBE) by enabling innovation and scientific discovery. However, little is known about SBE scientists' distributed collaboration, a vital practice that CI must support for the doing of science. To provide insight into this question we interviewed 12 scientists regarding their work practices as they pursue joint research projects with colleagues from other universities. We identify the most frequently used physical and digital tools for SBE science and collaboration and characterize commonplace scientific practices in this domain with a paradigmatic example.

Author Keywords
Distributed Collaboration, Documents, Scientific Practice, Cyberinfrastructure

ACM Classification Keywords
H.4 [**Information Systems Applications**] K.4.3. [**Computers and Society**]: Organizational Impacts – *computer-supported collaborative work* J.4 [**Computer Applications**]: Social and Behavioral Sciences – *sociology*

Introduction

We report the findings from a pilot study of twelve SBE researchers and identify personal digital assemblages they construct with their team in order to do the work of science. To do this, we ask the question: *What document infrastructures do SBE scientists build to support their virtual organizing and documenting practices?*

The rise of cyberinfrastructure (CI) is one of the defining features of modern science. For example, in domains such as climate change, high-energy physics, and ecology, the development of CI has served as a means for sharing large-scale technical resources, data, and intellectual skill specialties. The 2005 report published by the National Science Foundation (NSF) on cyberinfrastructure and the social sciences identified eight challenges and recommendations for moving forward with the development and deployment of CI: sharing data, advanced analysis, method sharing, and advancing intellectual domains of inquiry [1].

Our working premise is that these recommendations are derived from particular domains of natural and physical scientific practice and may not be relevant for research conducted in the domain of Social, Behavioral, and Economic sciences (SBE). Researchers in this area have different needs and practices, calling into question the appropriateness and efficacy of these guidelines for SBE researchers.

Approach

Scientific communities rely on the practice of documenting, inscribing and otherwise publishing findings as a form of validity, peer review, error checking, and disseminating knowledge [3]. A variety of physical and digital tools in differing configurations, thus materiality, are important factors in the doing of SBE science. Documents are central to the doing of science, and widely used digital tools such as email and Dropbox produce detailed log data and documentation of relevant work practices providing an opportunity for the study of these activities.

This perspective frames documents as key artifacts and products of work, especially scientific work, and – as such – are a window into scientist's work. Thus, documents serve as data to analyze because they encapsulate the flow of knowledge between actors, include demarcation of place and temporality, and serve as boundary objects for interdisciplinary collaboration [5][6]. Previous research has revealed that small-scale distributed science projects create personal assemblages of consumer or "off the shelf" solutions to go about [7].

Our sample consists of 12 academics -- 8 women and 4 men --, in the fields of the social (four), information (six), and computer sciences (two). 8 of these were pre-tenure, 2 were post-tenure, and 2 held post-doctoral or other research positions. All interviewed participants were at or connected to a research university. This sample of 12 participants comes from a larger pool of prospective participants who have been funded by national or corporate entities to conduct collaborative scientific projects.

Interviews were transcribed and content analysis was performed adhering to a constant comparison technique using both deductive and inductive approaches [4]. A set of detailed codes were developed to describe three primary concepts: information and

communication technology uses, documents, and project details. In addition to the interview, each participant completed a survey of what digital and non-digital tools and software they use for their work practices, the frequency with which they use each tool, and whether each tool is provided by the institution or is part of a personal account or personal use.

Findings

In this early work we have observed that much distributed SBE research is primarily short lived (2-3 year). Collaborations are phenomena-oriented, in that the focus of projects is on emergent or reoccurring social phenomena. Data collection for these projects occurs in the field, in situ, more often than in a laboratory environment. We have found that the tools used to facilitate distributed scientific collaborations are overwhelmingly commercial applications and software packages. These tools are carefully pieced together to create a personal infrastructure that provides solutions to problems such as sharing data, collaborative writing, synchronous and asynchronous communication, and data analysis.

Our analysis of the coded interviews reveals two types of collaborative approaches: either scientists have set ways of collaborating or they specifically look for tools and software that everyone in the group can use to create a streamlined ecosystem of compatibility. Table 1 shows some of the tools important to the doing of SBE science, as well as indicating whether those tools are primarily physical or digital. Many participants expressed a preference based on the material nature of tools, for example, taking field notes with pen and paper in contrast to staying "all digital" using a laptop or tablet to type notes. When coding for types of practices, we found an overwhelming amount of organizing done over email, video services like Skype, instant messenger, and face-to-face meetings. We collected a list of 58 unique tools used to support SBE work practice and collaboration. Table 2 shows the most popular software and tools that are *used daily or with a regular frequency*.

Table 1: **Tool Materiality**

Physical	**Digital**
Field Notes	R
Face-to-Face Meetings	Atlas.ti
Whiteboard	Microsoft Word
Paper notebook/journal	Dropbox
	Mendeley
	Google Calendar
	Skype

Three patterns of use emerge from the data on tools and software. First, a majority of the software is free or has a free option to use. Second, email is the preferred method of communication and, more specifically, every participant had a Google Gmail account in addition to any other email accounts they use. Finally, many of the investigators are connected to multiple Google products; this is in line with observations we have made about group members choosing software and tools based on the ability of each group member to use them. The Google ecosystem is one of the emergent personal infrastructures in SBE scientific collaboration that we observed in our preliminary sample.

To illustrate the typical document infrastructure that supports virtual organizing and remote collaboration of SBE scientific practices, we present a paradigmatic example. This case serves as an exemplar for the type of science we are studying and was selected by piecing together the practices that SBE scientists view as commonplace and standard [2].

Principal investigator, Jane Roe, collaborates with four other members on her small scale, short lived (2-3 year), and funded research project. Two of the collaborators are collocated at the same university with Jane. One is a post doc and the other a graduate student in the department. The other two collaborators

Table 2: **Commonly Used Tools**

Software/Tool	Users
MS Word	12
MS PowerPoint	8
Google Docs	11
Google Calendar	10
Google Mail	12
Google Drive	6
Dropbox	8
Endnote	4
Mendeley	5
Atlas.ti	4
R	4

are remotely located in other states; one is a co-principal investigator (co-PI) at another university, and the other collaborator is a subject matter expert with specific technical skills needed for the project. The group has weekly meetings in Jane's office, using a laptop with Skype to include the other two remotely located collaborators.

They are in the middle of their project timeframe, collaboratively writing papers for publication using a shared Google document; Microsoft Word is used for formatting and submitting the final article. The group uses Dropbox to share project files, a shared Mendeley group to manage project readings and citations, a collaborative Google calendar for the project, and they all communicate heavily and daily over email. While the group collaboratively writes a paper, the remotely located technical subject matter specialist uses the statistics package R for some of the data analysis, while the co-PI uses Atlas TI for the other portion of data analysis and transfers handwritten field notes to word documents to share with the group.

Next Steps

This work is part of an ongoing project to detail SBE scientists distributed collaboration practices. Moving forward we have identified key documents used by the participants in our sample. We are implementing methods for capturing document metadata and relevant project emails to shed light on specific SBE scientific practices. We plan to follow the documents that scientific collaborations produce and use, along with behavioral queries such as interviews, so that we may uncover the complex and highly personalized work practices that are common in SBE science.

Acknowledgements

We would like to thank our participants for generously sharing their time and experiences. This research is funded by the National Science Foundation (Award #1221945).

References

[1] Berman, F.D. and Brady, H.E. Final report: NSF SBE-CISE workshop on cyberinfrastructure and the social sciences. Arlington, VA, 2005.

[2] Flyvbjerg, B. Five Misunderstandings About Case-Study Research. Qualitative Inquiry 12, 2 (2006), 219–245.

[3] Lund, N.W. Document Theory. Annual Review of Information Science and Technology 43, (2005), 399–432.

[4] Muller, M.J. and Kogan, S. Grounded Theory Method in HCI and CSCW. 2010, 1–46. http://www.watson.ibm.com/cambridge/Technical_Reports/2010/TR2010.09 Grounded Theory Method in HCI and CSCW.pdf.

[5] Østerlund, C. and Boland, R. Document Cycles: Knowledge Flows in Heterogeneous Healthcare Information System Environments. Hawaii International Conference on System Sciences, (2009), 1–11.

[6] Østerlund, C. Documents in Place: Demarcating Places for Collaboration in Healthcare Settings. Computer Supported Cooperative Work (CSCW) 17, 2-3 (2007), 195–225.

[7] Sawyer, S., Kaziunas, E., and Østerlund, C. Social Scientists and Cyberinfrastructure: Insights from a Document Perspective. CSCW 2012, (2012), 931–934.