Documents and Distributed Scientific Collaboration

Matt Willis, Sarika Sharma, Jamie Snyder, Michelle Brown, Carsten Østerlund, Steve Sawyer

This paper asks 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 (SBE) sciences. CI enables innovation and scientific discovery. Little is known about SBE scientists' distributed collaboration. To provide insight into this area, twelve scientists were interviewed about their work practices as they work with colleagues from other universities.

The rise of CI is a defining feature of modern science. For example, in domains such as climate change, the development of CI has served as a means of sharing large-scale technical resources, data, and intellectual skill specialties. The National Science Foundation published a report on CI and social sciences that identifies eight challenges and recommendations for moving forward with the development of CI: sharing data, advanced analysis, method sharing, and advancing intellectual domains of inquiry. The working premise in this paper is that these recommendations are derived from particular domains of natural and physical science practice and may not be relevant for research conducted in the SBE field.

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. A variety of physical and digital tools are important factors in 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. This perspective frames documents as key artifacts and products of work, especially scientific work, and as such are a window into scientist's work.

In the early findings the team observed that much distributed SBE research is short lived. Collaborations are phenomena-oriented, research focuses on emerging or recurrent social phenomena. Data collection for these fields occurs in the field more than in laboratory environments. The tools used to complete work is overwhelmingly commercial applications and software packages. The analysis of interview reveal two types of collaborative approaches: either scientists have set ways of collaborating or they specifically look for tools and software that everyone on the group can use to create a streamlined ecosystem of compatibility.

When coding types of practices, it was found that an overwhelming amount of organizing was done over email, video services, instant messenger, and face-to-face meetings. Three patterns emerge from the data on tools and software. First, a majority of the software is free or has a free option. Second, email is the prefered method of communication and every participant has a gmail account in addition to

other email accounts. Third, many of the investigators are connected to multiple Google products. The Google ecosystem is one of the emergent personal infrastructures in SBE scientific collaboration.

This work is part of an ongoing project to detail SBE scientists distributed collaboration practices. Moving forward the team has identified key documents used by the participants in the sample. They are implementing methods for capturing document metadata and relevant project emails to shed light on SBE scientific practices. The plan is to follow the documents that scientific collaborations use and produce, to undercover the complex and highly personalized practices that are common in SBE science.