The Use of Email in Distributed Scientific Collaborations[1]

Matt Willis; Carsten Oesterlund; Sarika Sharma; Nick Kaptyn; Steve Sawyer
{mwillis; costerlu; skshar01; nkapteyn; ssawyer} {@syr.edu}
School of information Studies, Syracuse University, Syracuse NY 13224,

Project info @ http://voss.syr.edu

This project update comes to share some of what we are learning about email uses in distributed scientific collaborations. We also share out the visualization strategies we have developed to assist our email data analysis.

As might be expected, email serves as a – if not *the* – primary mechanism for interaction, communication and information sharing among distributed scientific collaborators. While collaborators come from a range of disciplines, the 22 projects we have studied are grounded in social science practices around similar research interests, providing a basis for comparison. Analysis highlights three distinct roles email plays in supporting distributed scientific collaboration: (1) articulation, delegation and coordination work; (2) document management and archiving; and (3) shared cognition.

First, much of the email among scientific collaborators focuses on either articulation work – work done in order to do goal-oriented work (e.g., following the right template in order to submit a paper) – or on project goals, meeting times, detailing tasks and to-dos, and coordinating schedules. Second, while document management and file sharing software/platforms are widely available, we find file sharing through email is a common practice in these distributed collaborations. Email appears to serve as a redundant file of record. Team members often put a file in a shared repository such as Dropbox, Google Drive, or a university specific repository. However, team members also attach the same file to an email to edit and review. Third, we find email serving as social cognition. It is a script or record for what has happened previously for the project, and helps plan the future goals and tasks of the project. It is also visible, arriving to each participant's email inbox to be seen by each member but each member can work with the email separate from the rest of the collaboration.

Findings make clear that email practices are deeply integrated with the lives of the scientists we studied. Email messages often contain personal or non-project related information, come in waves, and email practices and etiquette shift over time in a single group. One shifting practice we observe is that email subjects do not always relate to or represent the content of the email body. This issue of subject field context presents a challenge for studies that collect email metadata and rely on subject lines to derive context of the email. We conclude that project emails are the distributed digital equivalents of physical laboratory notebooks and, as such should be used as part of the team's digital interactions.

Beyond understanding the role and practices of email use in collaborations, we want to understand the structure and flow of email. To do this we use qualitative content analysis of individual emails followed by use of visualizations. Emails are tagged based on inductively coded themes. Two primary parent codes are prevalent in the emails: (1) *scientific activities*, which consist of: funding, grant writing, publications, literature review and data analysis. The second parent code, (2) *admin. efforts*, represents

[1] This work is supported by US NSF grants 1221945. Any opinions, findings, conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the NSF.

concepts like: meeting, coordination, discussion, and sharing work.

After coding we developed a custom script for the *R* statistics package to help visualize this data. Meta data of use includes sender, receiver, attachment name, attachment number and tags for scientific activities and email genre. For example, Figure 1 displays the team member on the y-axis and the date on the x-axis. This figure represents the bursts of email that happens by each team members in the month of September 2014. Each dot represents an email and size of the dot indicates frequency. Red dots indicate an attachment. Team members are randomly positioned.

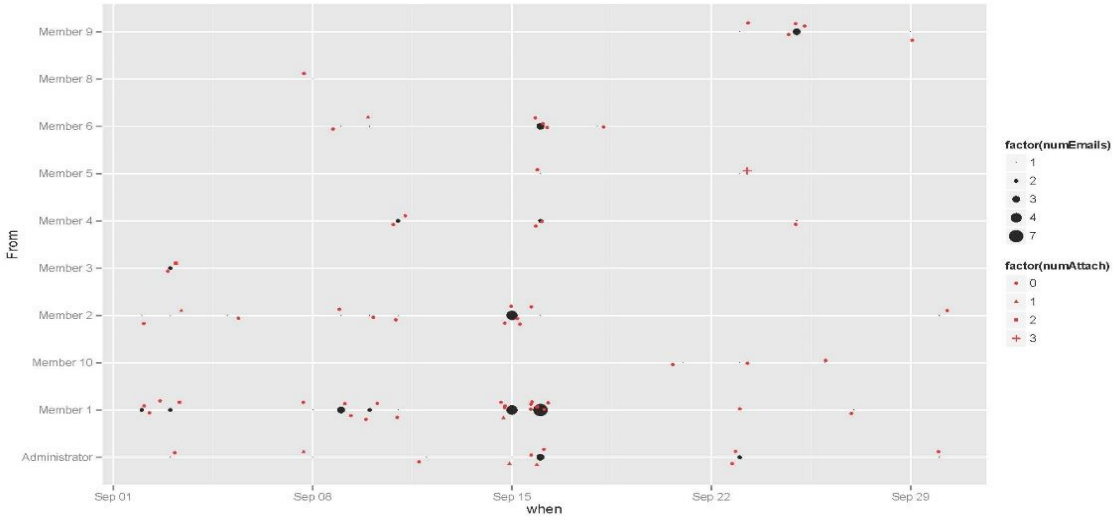

Figure 1: September 2014 email communication of a distributed scientific collaboration

Figure 2 displays team members on the y-axis and the scientific activity associated with the email on the x-axis. Each black dot represents an email and size indicates frequency. Red dots indicate an attachment and each red symbol indicates frequency. This graph displays the bursts of scientific activity that took place in the month of September 2014.

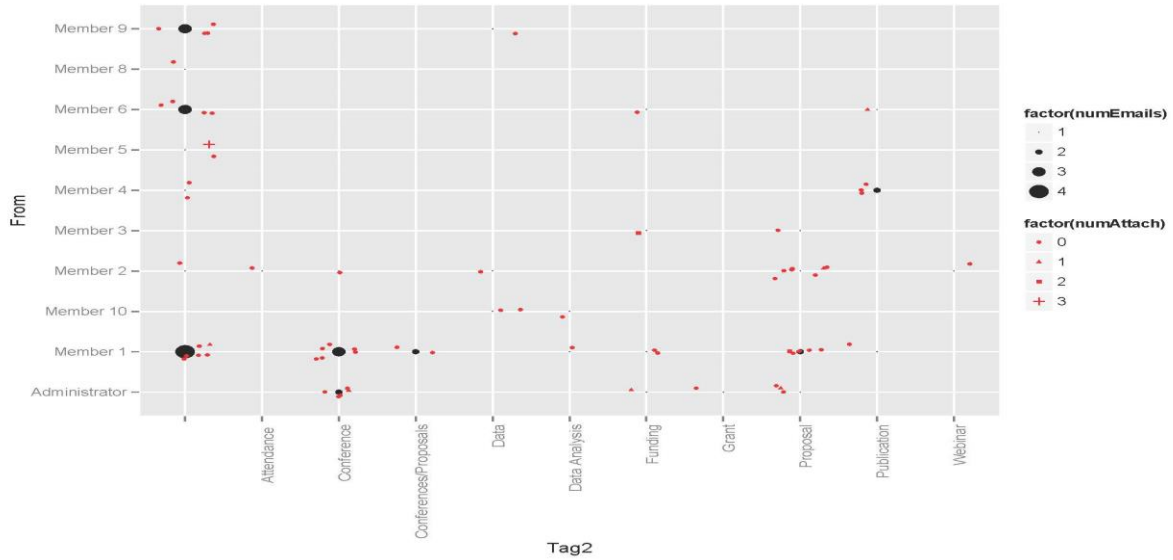

Figure 2: September 2014 email associated with scientific practices

We have found visualizations to serve as a pictorial representation of collaborative work that unfolds asynchronously over a long period of time through many email messages. The visualization graphs are also used as an elicitation tool to further interrogate in greater detail the collaborative work that takes place on email. Participants rarely think about their collaboration in terms of temporal email rhythms. Presenting participants with a graph of longitudinal email provides a clear sense of the temporality and flow of email. Using these visual representations of email flow and temporality participants can see increases in email around conferences, publication writing, and other deadlines. We then ask participants to unpack these events that have increased email exchanges with the goal of understanding the articulation work and creation of documents that must be done for the collaboration to hit the deadline. Email is ubiquitous and transparent in the collaborations we have studied. Methodological and analytical techniques, such as graphing, that make email practices visible may shed new light on the role of email in collaborations. Our preliminary understanding on the use of email based on our findings is that it functions as a continuously edited document infrastructure that is required to produce other project documents, publications, presentations, written communication, and in general: do the science of the collaboration.

Please feel free to cite this as:

M. Willis, C. Oesterlund, S. Sharma, N. Kaptyn and S. Sawyer (2015), The Use of Email in Distributed Scientific Collaborations, Project Memo #2, School of Information Studies, Syracuse University, October 15, 2015, http://voss.syr.edu/publications/VOSS-email-blast2.pdf.

For other publications you may be interested in, please see:

Sharma, S., Snyder, J., Østerlund, C., Willis, M., Sawyer, S., Brown, M., & Skolozar, D. (2014). Document practice as insight to digital infrastructures of distributed, collaborative social scientists. In iConference (pp. 1021 – 1024). Berlin, Germany.

Willis, M., Sharma, S., Snyder, J., Brown, M., Østerlund, C., & Sawyer, S. (2014). Documents & distributed scientific collaboration. In CSCW. Baltimore, MD.