Sociomaterial Processes, Long Term Planning, and Infrastructure Funding: Towards Effective Collaboration and Collaboration Tools for Visual and Data Analytics

Kerk F. Kee, Ph.D.
Assistant Professor
Department of Communication Studies
Chapman University
Orange, CA 92866
kerk.kee@gmail.com

Larry D. Browning, Ph.D.
Professor
Department of Communication Studies
University of Texas at Austin
Austin, TX 78712-0115
lbrowning@mail.utexas.edu

Dawna I. Ballard, Ph.D.
Associate Professor
Department of Communication Studies
University of Texas at Austin
Austin, TX 78712-0115
diballard@mail.utexas.edu

Emily B. Cicchini, M.F.A.
Special Projects Manager
Instructional Technology Services
College of Liberal Arts, UT Austin
Austin, TX 78712
emilyballcicchini@utexas.edu

"Science of interaction" in the context of data and visual analytic systems involves human-human communication (i.e., collaboration over data) and human-data interaction (i.e., collaboration with data) as a complete computational-human decision making loop. At both levels, this developing science of interaction challenges traditional notions of time—especially in terms of pace, sequence, and planning (Ballard & Seibold, 2003, 2004). The capacity of the visual analytics themselves affords researchers a much shorter path from discovery to presentation of ideas in their collaboration *with* data. The implications of this truncated sequence is one aspect of the process that we address in this position paper. As well, researchers have recently addressed the need to consider The Long Now of Cyberinfrastructure (Ribes & Finholt, 2009). The implications of this Long Now for collaboration *over* data involving visual analytics is the second aspect of the complete computational-human decision making loop that we address in this position paper. Particularly, while a complex web of human agents and large-scale data (Bird, Jones, & Kee, 2009; Kee, Cradduck, Blodgett, & Olwan, 2011) are the key ingredients in this decision making process, we focus on the time involved in the *development* of cyberinfrastructure (CI) tools and the *implementation* of these tools to enable data-intensive and visual analytic collaboration. By extracting the narrative rationality(ies) (Browning, 1992; Browning & Boudes, 2005) of an existing dataset, we highlight challenges that deserve research attention and then provide recommendations for agenda setting.

Planning for Effective Collaboration and Collaboration Tools Development

<u>The Lack of Capability Imagination by Scientists</u> –CI tools development for visual analytic and collaborative interactions is driven by scientists' research questions and needs. But scientists often have difficulty envisioning the science—therefore their needs—that could be served by these tools.

> "We're at the point of prototyping and developing what might be possible. That process of working with NCSA [National Center for Supercomputing Applications] and other organizations to define what cyberinfrastructure should be and to prototype early cyberinfrastructure…We're trying to build a national virtual community with WATERS Network… How you bring together these really diverse communities and get them to work together effectively, we struggle with that… The virtual organization cyberinfrastructure has a long way to go before it's going to meet the needs of the scientific communities that are trying to do these large scale collaborations. Part of the problem is we don't even know what we want… Well, that's the problem. It's hard to come up with what it should look like when we've never had it before" (Water Resources Engineering Researcher, Illinois).

<u>The Lack of Domain Knowledge by Technologists</u> – CI tools are often co-produced by domain scientists (with domain knowledge, organizational practices, disciplinary norms, etc.) and technologists (with computational expertise, distributed and parallel programming skills, etc.). However, the tools developed cannot be effectively and efficiently built without substantial knowledge of the scientific domains and disciplinary practices on the part of technologists.

> "Most of these projects struggle with this… Many of them struggle with the fact that the programmers that you hire, any IT researchers, need to learn the science that is being served through these systems… When we started building a registry for geological samples to give unique identifiers to the sample, talked for a long time about the various sampling techniques and the sample types that you can get and so on to make the system as generic as possible. After two hours of discussion about

different sampling locations and it could be an outcrop, it could be a drill core and so on, finally someone from the IT group asked – So what is an outcrop? Very fundamental, extremely basic concepts of geology were not known to them. You had to go back many times and iterate on the system requirements because their knowledge kept growing and they said – Oh, if this is that way, then we cannot do it here. It causes a substantial delay, I think, from every side – from my side, and from the side of the IT developers. We were very unrealistic about the time frames in doing this development, because they did not consider the learning curve" (Geochemist, New York).

Design of Collaboration and Tools Implementation

The Challenge of File Formats that Slows Down Visualization/Modeling Result Sharing - The process of visual/data analytics and collaborative interactions sometimes can be hampered due to modeling results in file formats that cannot be shared instantly in traditional presentation forms.

> "I use Linux for my platform to run my modeling. And my results are not a .doc or a powerpoint or a PDF. And this only allows certain extensions when you upload the data. So that's one restriction that we have that just hasn't been changed yet… I tried putting some modeling results up here and it didn't let me. So that was a frustration… We're going to have to hit bumps and have to go smooth them over as we develop these things" (Water Resources Engineering Researcher, Texas).

The Challenge of Combinational Integration between Academic and Commercial Technologies – Scientific collaborations are not driven simply by analytical tools alone, but also by communication tools that enable productive interactions among scientists. However, at this point, scientists appear to use multiple disconnected tools to facilitate collaborations. The design of such collaborations may need to explore ways to smoothly integrate both academic tools and commercial tools that scientists adopt to communicate with each other.

> "[T]he tools we use are very disconnected. So you have your tools you do your analysis with. Then when you've got results, you maybe share them with a colleague through something like WebEx or an on-line repository where you share the file. But you're not actually engaging with them remotely during the research process. So if someone comes over and meets with me, we can sketch things out and we interact in a different way than when we're not there together. We haven't found a good substitute for how to do that remotely" (Water Resources Engineering Researcher, Illinois)

Balancing Face-to-Face, Asynchronous Mediated, and Synchronous Mediated Communication – While NSF's Office of Cyberinfrastructure is enthusiastic about the potential of building effective CI and virtual organizations that support large-scale scientific collaborations, a strategic mix of various communication forms may be necessary at this point to enable the most impactful collaborations.

> "In 2005… ISI, Science Indicators called us up and told us that we've been ranked as number 22 in the world out of 38,000 institutions in terms of our scientific impact… This kind of research is clearly needed and the approach that we use, bringing various experts into NCEAS [National Center for Ecological Analysis and Synthesis] for relatively short periods of time when… it's sort of a high productivity, high energy time … [I]t's like a hybrid… Some of the collaborative infrastructure that is available now is making those times in between personal meetings more productive" (Biologist, California).

Recommendations for the Research Agenda of a "Science of interaction" for Data and Visual Analytic Systems:

1. Examine the sociomaterial processes (Leonardi, 2009, 2010) of going directly from findings to presentations (i.e., using Data Visualization tools, PowerPoint, Acrobat, etc.) such that direct findings are in (dispersed and collocated) presentations—and sometimes dynamically presented to show variable flow and process of interaction. Identify the effect of findings going directly from research to presentation and the forms the presentation takes, without any reformulation on a variety of cooperative effects—e.g., understanding, use in future research, etc.—in data and visual analytics. This sociomaterial understanding would shed light on how visual analytics involves processes "from individual manipulation of data representation, to interactive cognitive discovery combined with automated analysis, to coordinative and collective interactive analysis among groups of individuals", as stated in the call. This understanding will also attend to the challenges raised in the narratives provided in the "Design of Collaboration and Tools Implementation" section.
2. Focus on long term planning (and training/education) in order to avoid lack of domain knowledge, lack of interoperability standards, disconnected analytic tools and communication media, etc. If we envision ourselves in a long term project, we will work on planning at the start. If we envision ourselves in a project that must launch next week, we will not. While progress has been made that provides opportunities to test models using large sets of shared data, especially through data management plans, the field needs to be planned for the long term. One outgrowth of this approach will be to move solely

from Weick's (1979) notion of communication cycles (i.e., developing ever-new responses and reactions to emerging problems associated with a given virtual organization) to developing assembly rules (i.e., more standardized operating procedures—specific forms, formats, and structures—and routines for data and visual analytics from which multiple virtual organizations may benefit). This approach will attend to the narrative challenges raised in the "Planning for Effective Collaboration and Collaboration Tools Development" section of this position paper.

3. Advocate for long term funding commitment from NSF (Kee & Browning, 2010) to sustain such an effort, instead of the usual 3-year and 5-year funding cycles for selectively funded projects, so that participating scientists and technologists will have more successful collaborations in the computational-human decision making process as well as effective tool development. Focus on *infrastructure time* instead of *project time* (Karasti, Baker, & Millerand, 2010). This effort will attend to all the challenges raised in the narratives presented in this paper, balancing the tensions between "project or facility, development or maintenance, and inclusion or innovation" (Ribes & Finholt, 2009, p. 6) as well as taking steps toward an emergent future (Karasti, et al., 2010) of effective collaboration and collaboration tools for visual and data analytics.

References

Ballard, D. I., & Seibold, D. R. (2003). Communicating and organizing in time. *Management Communication Quarterly, 16*(3), 380.

Ballard, D. I., & Seibold, D. R. (2004). Organizational members' communication and temporal experience: Scale development and validation. *Communication Research, 31*(2), 135-172.

Bird, I., Jones, B., & Kee, K. F. (2009). The organization and management of grid infrastructures. *Computer, 42*(1), 36-46.

Browning, L. D. (1992). Lists and stories as organizational communication. *Communication Theory, 2*, 281-302.

Browning, L. D., & Boudes, T. (2005). The use of narrative theory to understand and respond to complexity: A comparative analysis of the Cynefin and Weickian models. *Emergence: Complexity & Organization*, *7*(3), 32-39.

Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work, 19*(3), 377-415.

Kee, K. F., & Browning, L. D. (2010). The dialectical tensions in the funding infrastructure of cyberinfrastructure. *Computer Supported Cooperative Work 19*(3), 283-308.

Kee, K. F., Cradduck, L., Blodgett, B., & Olwan, R. (2011). Cyberinfrastructure inside out: Definitions and influences shaping its emergence, development, and implementation In D. Araya, Y. Breindl & T. Houghton (Eds.), *Nexus: New intersections in Internet research.* (pp. 157-189). New York: Peter Lang.

Leonardi, P. M. (2009). Crossing the implementation line: The mutual constitution of technology and organizing across development and use activities. *Communication Theory, 19*, 278-310.

Leonardi, P. M. (2010). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly, Forthcoming*.

Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems, 10*(5), 375-398.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). Reading, MA: Addision-Wesley.