CASE STUDY 1: E-SCIENCE AND AWARENESS

The focus of this case study is on the exploitation of "awareness" capacity in the mediated environment of e-science and one particular instantiation within the e-social sciences.

"Awareness": Conception on top of Content

The primary focus of much of the e-science literature is on the use of CSCW systems for the consumption and production of scientific knowledge, where the technological infrastructure acts only as a geographic and temporal mediator between actors in the shared knowledge space. Such a model, however, ignores one of the critically unique byproducts of such a mediating framework: unlike unmediated face-to-face communication, the entire process of communication may be recorded and archived, not just the final output of the interaction. For example, in a traditional collaborative endeavor, a group of scholars at a particular facility might all meet in a conference room and discuss their latest findings, collaboratively constructing a final report that is eventually submitted for publication. In the electronically-mediated version of this meeting, however, correspondence occurs through email, instant messaging, and automatically transcribed video conferencing such that every interaction, no matter how slight (an email asking whether someone will have a particular paragraph done by that afternoon) becomes part of the permanent organizational record of the execution process that then becomes attached to the final output of that episodic collaboration. This "provenance package" offers both an attribution record associating each component of the report with all of the actors that shaped its evolution, as well as a "recipe" to guide others hoping to repeat the process for other projects at a later time.

The technological tracking and utilization of such user actions is known as "awareness" and has become an increasingly important aspect of e-science environments. The ability to coordinate activities across a workforce that transcends geographic and temporal boundaries requires continual knowledge of what collaborators are doing and saying and of the unfolding events of the work environment. A key facilitator of this process is *communicative awareness*, the knowledge of communication patterns in a collaborative team. As argued by Gutwin, Gwen, and Greenberg, (2004) "*awareness* of other group members is a critical building block in the construct of team cognition" which leads to the conclusion that "*computational support for awareness* in groupware systems is crucial for supporting team cognition in distributed groups."

Social Translucence

How exactly is the notion of "awareness" in CMC systems, such as those used in e-science, understood and described in the literature? Erickson & Kellogg (2000) introduce the notion of *socially translucent systems* that offer technologically-mediated awareness mechanisms to coordinate cooperative action within a CMC environment. They offer the

example of a windowless door in their office building that opens from the stairwell into the hallway. The door is often opened quickly as people exit the stairs, slamming into anyone unlucky enough to be standing on the other side at that moment. The company's solution to the problem was simply to post a sign asking passerby's to "Please Open Slowly," which, as expected, did little to stem the problem. Without any method of seeing what lies beyond the door, coupled with the low likelihood that someone will be behind it at any given moment, the average person will swing it open in cavalier fashion. Adding a small window to the door, on the other hand, alleviates the entire issue, as it offers parties on both sides of the door the ability to determine what lies beyond and modify their behavior accordingly.

The authors describe the outcome of this *social translucence* in terms of *visibility*, *awareness*, and *accountability*. While both sign and window convey the same potential for a collision, the window makes that potential more visible by exposing it in a more intuitive (and therefore recognizable) fashion. Potential does not equate to actuality, and the low likelihood that any given opening of the door will result in someone being hit leads the sign to have little realized impact. A window, on the other hand, represents a *ground truth*, definitively warning that someone is on the other side. Eliminating false positives means that warnings are more likely to be taken seriously. Finally, even if the person on the hallway side of the door does not see the person in the stairway opening the door and hitting him, there is a level of accountability to the opener, because both parties know that the opener had both the opportunity and expressed expectation to utilize this capability to prevent such an occurrence. Mutual knowledge of the shared portal therefore reinforces its use, as the consequences for not doing so are increased.

Within the e-science "collaboratory" model, progress has been made towards capturing basic levels of awareness, but as Erickson & Kellogg note, the current generation of systems still represent a relative infancy in awareness technologies. Systems often tend to focus on locative and availability indicators, offering actors a list of others currently signed into the system. Those that do offer structural-based awareness connectivity tend to do so only along bibliometric dimensions associated with the final data, as in VisBench's ability to suggest rendering algorithms and configuration settings based on those using similar datasets. (Heiland, 2001)

Awareness in E-Social Science

E-social science refers to utilization of technological mediators in the conduct of social science research. The field draws many supporting parallels from the more entrenched traditional e-science discipline, but must contend with a uniquely human-centered and human-centric approach to work execution. In the traditional sciences, a majority of the work is conducted using the support of specialized equipment or software environments which perform the execution components, while the human actors perform the analytical synthesis of the results. In the e-social sciences, however, human knowledge workers comprise the primary execution engine of the research pipeline, underlying every step of the process, from data collection, to cleaning and preparation, to computation, to final

presentation. The enormous interlinking web of humans that underlie large e-social science projects necessitates large teams with correspondingly large management requirements. At the Cline Center for the Democracy at the University of Illinois, the Comparative Constitutions Project is one such example of a new breed of e-social science endeavor, an extremely comprehensive, computer aided, review of every written constitution for all the nations of the world from 1789 to present. The massive scale of the project necessitates a large distributed team of nearly 30 students and faculty that manage the considerable workflow of acquiring, reading, and coding more than 3,000 constitutional events using their computer-based coding protocol. A web-based asynchronous message board system facilitates communication among the project's active team members, coordinating the activities of its distributed workforce using a collection of role-specific discussion forums.

When the constitutions project first began using the messaging system, one of the chief concerns of the faculty were that they were struggling to understand the communicative patterns of their forums. While they could browse through each forum's collection of postings, faculty wanted the ability to see a high-level view of general patterns, such as which students had the highest posting volume and who the faculty were particularly good about replying to quickly, along with more advanced indicators, such as topical trends over time, and the structure of the informal communications networks students established among themselves.

An asynchronous message board contains a dichotomy of structured (in the form of message flows) and unstructured (in the form of message content) data which yields significant, yet very different, insights into its communicative activity. Each message is associated with a sender, a recipient (the particular message board, forum, and thread it was posted to), and the date it was posted. Messages which are replies to other messages also contain information about the message and thread they were posted in reply to. In addition, various derived structural metrics such as first reply latency and an individual user's ratio of new threads to replies, may be computed. Statistical analysis techniques and social network theory may be readily applied to such structural data to yield quantifiable insights into the communicative patterns of the message board. Similarly, each message's body text contains sequences of nouns, called "noun phrases," which may be used to approximate the topical content of the message. Using these techniques, the freeform text of each message is transformed into a set of phrases, or semantic surrogates, describing key concepts in the message, which may then be used to relate messages together at the content level.

The message board system at the Cline Center packages all of these awareness indicators into an advanced suite of integrated inline and external realtime displays that allow faculty to maintain an up-to-the-moment view of the communicative patterns within their community. Prior to the deployment of this awareness framework, the constitutions group had weekly face-to-face meetings with all onsite personnel meeting in a crowded conference room and offsite personnel calling in via conference call. Faculty would go around the room, asking each student in turn to briefly outline her progress, outstanding problems she was facing, major questions, etc. The limited amount of time of the

meetings, however, severely restricted the amount of information faculty were able to gather each week, preventing them from really delving into the deeper patterns and status of the project. Within a month of changing over to the new message board system with its new awareness measures, faculty were able to completely eliminate the weekly meetings, a very strong indicator of the success of such mediated awareness indicators.

In the case of this project, the *content* of the message boards became less important than the analytical tools that made it possible to explore the *creation* of that content and the patterns of creation and collaboration that underlay each final board posting. This presents a unique notion of the creation process as described through such awareness indicators as a form of "knowledge" in the distributed knowledge workflow of e-science. Rather than "knowledge" being only the final output of the collaborative process, the term is modified to encapsulate the process that led to the production of that knowledge, offering a new conceptualization of distributed knowledge production in the e-sciences as *e-knowledge = creation + content.*

Looking to the Future

As we examine today's state of the art and look to the future of e-science systems, we must ask ourselves what lies beyond the next horizon in terms of how awareness will be incorporated into the next generation of e-science collaboratory portals? In particular, to what degree can systems learn enough about the conversations they mediate and the actors they support such that they may begin to take over for those actors in specific circumstances, acting as intelligent agents? As more and more information continues to be made available in electronic format and as enterprises continue to launch mass digitization endeavors to bring their paper records into the digital era, knowledge workers are faced with an ever-expanding universe they must sift through to find the slide of information they need. Couple this with the vast amount of additional archival data that awareness measures make possible and the fact that most large corporations now use enterprise content management systems that archive every document, email, and other form of communication for access, it becomes readily apparent that new ways of penetrating this fog of information will become necessary.

To what degree can technologically-mediated awareness help with this information overload condition? Already, prototype search systems leverage the "wisdom of the crowds" by allowing users to publish bookmarks of their favorite links for a particular topic, or by automatically mining the search patterns of hundreds of millions of users to dynamically rank the order of search results based on the selections of others. The degree to which these consumer-oriented technologies can find their way into e-science collaboratories will have a significant impact on whether such collaborative environments can expand beyond the small project-specific portals of today into the broader discipline-wide environments of tomorrow.

References

Erickson, T., & Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction, 7* (1), 59-83.

Gutwin, C., Gwen, S., & Greenberg, S., (1995) Support for Workspace Awareness in Educational Groupware, In Proceedings of *CSCL'95 Conference on Computer Support for Collaborative Learning*, Bloomington, Indiana, pp 147-156.

Heiland, Randy W.; Baker, Pauline M.; Tafti, Danesh K. (2001). VisBench: A framework for remote data visualization and analysis. *ICCS 2001, San Francisco, CA.* v. 2074, pp 718-727.