The Changing Face of Digital Science: New Practices in Scientific Collaborations

NSF Workshop on Discovery Informatics, January 2012

Cecilia Aragon

Department of Human Centered Design & Engineering, University of Washington

Introduction

As computation has become a fundamental tool for the scientific method in recent years, there has been a concomitant rise in scientific collaborations of both collocated and geographically distributed teams. This trend towards collaboration introduced a need for supporting technologies, and scientists were among the first to adopt information and communication technologies into their work practices. Tools such as wikis, instant messaging, and email have been widely adopted to aid in collaborative efforts [1, 2].

Today, a new generation of physicists, biologists, and other scientists is changing the course of scientific research. Scientists who grew up with Facebook, Twitter, and IM are developing and applying new means of collaborating to the scientific process. Scientists are beginning to explore the use of social networking sites, blogging, and microblogging for information exchange. Existing social tools such as chat, IM, and FriendFind are being adopted and modified for use as group problem-solving facilities.

At the same time, the landscape of science itself is shifting. Increasingly, scientific research is conducted by large, multi-institution and interdisciplinary project teams, processing exponentially vaster and more complex data flows. This overwhelming increase in the amount of scientific data being generated has been called the "data tsunami." The inability to gain insight into complex scientific phenomena using current software tools is a bottleneck facing virtually all endeavors of science, because even as data volume is increasing exponentially, human cognitive capacity is remaining relatively constant.

The confluence of the growing availability of social software with the increasing need for scientific collaboration to generate, analyze, and derive insight from vast and complex data sets is leading to a number of interesting developments.

Scientific Social Data Analysis

A new class of Web site has recently emerged that enables users to upload and collectively analyze many types of data (e.g., Many Eyes [3] and Swivel [4]). These are part of a broad phenomenon that has been called "social data analysis". This trend is expanding to the scientific domain where a number of collaboratories are under development. As the cost of hardware decreases over time, the cost of people goes up as analyses get more involved, larger groups need to collaborate, and the volume of data manipulated increases. Science collaboratories aim to bridge this gap by allowing scientists to share, re-use and refine their computational tasks (workflows).

To analyze and understand scientific data, complex computational processes need to be assembled and insightful visualizations need to be generated, often requiring the combination of loosely coupled resources, specialized libraries, and grid and Web services. The heterogeneity of the data, its size, and location, greatly complicate the data analysis pipelines. What are the requirements for the successful

development of scientific social data analysis software? Can such tools, e.g. VisTrails [5], facilitate scientific insight?

Barriers to the Adoption of New Collaboration Technologies

Although it was once predicted that scientists would "lead the way in making boundaries of distance obsolete and would be the first to take advantage of new technologies to assemble larger-scale efforts across distance," barriers exist that make scientific collaboration difficult [6]. Studies have shown that adoption of technologies can be hindered when they do not complement or are incompatible with existing work practices [7].

An issue of trust arises when data is made available on the web, since the web can "by-pass many of the social and technical processes by which communities decide what is known, what is to be trusted, what is accepted as public, published information."[8] The appropriation of social software for information exchange further complicates the questions regarding the reliability and security of information exchange.

In addition, although many technologies make sharing of information increasingly simple, knowledge is still difficult to transfer [9]. Knowledge is often difficult to represent and changes rapidly, but common understanding can be negotiated. What role can social software play in developing common ground in relation to knowledge artifacts?

Summary

How is the confluence of these two major trends (generational turnover in the scientific field and the oncoming data tsunami) impacting science and the process of scientific collaboration? How can HCI research elucidate the nature of scientific collaboration infrastructure? Can the study of human cognitive limitations provide insight into the difficulties facing scientific collaborations? Are there new approaches to scientific collaboration software that are proving successful in addressing scientists' challenges?

References

1. "The Changing Face of Digital Science," C. Aragon, S. Poon, and C. Silva, CHI 2009 Workshop Extended Abstract: ACM Conference on Human Factors in Computing Systems (2009).
2. Sonnenwald, D. 2007. Scientific Collaboration: A Synthesis of Collaborations and Strategies. In Cronin, B. (ed.), Annual Review of Information Science and Technology, vol. 4. Medford, NJ: Information Today.
3. Many Eyes, http://services.alphaworks.ibm.com/manyeyes/home/.
4. Swivel, http://www.swivel.com.
5. The VisTrails Project, http://www.vistrails.org.
6. Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., et al. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. Journal of Computer-Mediated Communication, 12(2), article 16.
7. Duque, R.B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.S., & Shrum W. (2005). Collaboration paradox: Scientific productivity, the internet, and problems of research in developing areas. Social Studies of Science, 35(5), 755-785.
8. Nancy A. Van House, Mark H. Butler, Lisa R. Schiff (1998). Cooperative Knowledge Work and Practices of Trust: Sharing Environmental Planning Data Sets. CSCW '98: The ACM Conference on Computer Supported Cooperative Work. Proceedings. Seattle, WA, November 14-18, 1998. ACM, 1998, p. 335-343.
9. Szulanski, G. (1992). Sticky Knowledge: Barriers to Knowing in the Firm. London: Sage Publications.