Looking Inside the Black Box: Computational Tool Development in Big Data Science

Len Hamilton, Whitney Caldwell, and Kerk Kee
COM-491: Spring 2014, Chapman University; Orange, California

Introduction

The term *black box* is technically and figuratively a device or system that is described solely in terms of its inputs and outputs (Winner, 1993). Winner (1993) elaborates on the metaphor that, "One need not understand anything about what goes on inside such black boxes. One simply brackets them as instruments that perform certain valuable functions" (p. 365). However, Leonardi (2009) explains, the social constructivist position (e.g., Pinch & Bijker, 1984) calls for discovering what happens within those black boxes as it is critical for understanding the dynamic between technological and organizational change.

Similarly, successful tool development in interdisciplinary teams is a *black box* in big data science. Big data is an emerging phenomenon, including science and engineering research, where many software tools are still in experimental stages. An understanding of the interactions between domain scientists (as users) and computational technologists (as developers) is needed for increasing the successes of these prototype tools. Due to the distribution of scientists geographically and more importantly, intellectually – there are critical barriers to achieving cohesive collaboration and effective communication throughout the development and implementation processes of these tools.

Literature Review

While many studies look at static innovations (such as mass produced or commercial technologies), this project examines computational tools as dynamic innovations (products that are user-driven and custom-made). In order to better understand these dynamic innovations and adjust our view on technological/organizational change to mutually constitutive, researchers must explore the relationship between three sets of activities (Leonardi, 2009), two of which include discovering what activities occur within the pioneering user/developer group during tool *development* and *implementation*. As the professional diversity of collaboration increases, the communication required to support the collaboration becomes more complex and challenging (Cramton, 2001). During the development and implementation stages there are key activities that facilitate the production of a successful tool. Thus, this poster seeks to answer the research question, "*What key activities constitute productive collaborations among multidisciplinary participants in computational tool development and implementation in big data science?*"

Methodology

This poster employed the grounded theory approach (Corbin & Strauss, 1990) and analyzed 25 interviews conducted with domain scientists (in bioinformatics, computational chemistry, theoretical physics, etc.) and computational technologists. Interview participants came from across the US (including CA, IL, IN, SC, MI, TX, etc.) and three from the UK. Interviews ranged from 16 minutes to 2:25 hours (10 conducted in person at the Supercomputing 2013 conference in Denver, and 15 over the phone, between Nov 2013 and April 2014). Guided by the stated research question, the co-authors performed multiple iterations of data analysis and literature integration, yielding preliminary findings presented in this poster.

Findings

Maintaining a "Bridge" Person

The first activity is to appoint and maintain a "bridge" person as an individual who can close the gap that exists between computational technologists and domain scientists. Oftentimes there needs to be someone that knows enough about both fields to be able to know which questions to ask, how to ask them, and can facilitate discussion between these two differing fields.

- "So in the most successful e-science projects I've seen, there's always been this sense of someone that I call a research facilitator, so it's someone who is bridging the gap." (Institute Administrator/ Theoretical Physicist, UK, 11/18/13)
- "…my goal my purpose, the role I was playing was I was trying to teach the computer scientists the domain aspect,… I always played that role; that bridge role." (Program Manager for XSEDE, IN, 11/18/13)
- "…it's very important to have a person whose job, we call it project managers, whose job it is to make sure that everyone who is in the virtual organization is working." (Center Administrator, CA, 11/21/13)

Developing a Common Language

In projects involving diverse domain scientists and computational technologists, participants must develop a common vocabulary that is understood and shared by both fields because domain scientist often don't understand the jargons of computational technologists and vice versa

- "An understanding of how to translate between vocabularies is the other important one because it's quite difficult." (Institute Administrator/Theoretical Physicist, UK,11/18/13)
- "…so you're talking about enabling vocabularies that allow software developers, managers, outreach trainers, and scientists, students all to cooperate and articulate where they think issues are and articulate what they think next stages should be together." (Institute Administrator/Theoretical Physicist, UK, 11/18/13)
- "…even language, there is a little bit of a language barrier in terms of understanding that since we don't have immediate audio/video, I can't see expressions, I can't see hand gestures, so when there is a language barrier all these extra clues that you get of what is happening go missing." (Computational Molecular Biologist, LA, 4/22/14)

Promoting a Cohesive Goal

In a project that involves multiple fields of discipline, the team must continuously promote a common goal that everyone strides for. The individuals in their respective field of study must understand that they alone cannot complete the project and solve the scientific problem. In order for the project to succeed they must put their benefits aside and work interdisciplinary to collaborate and solve the grand challenge.

- "…you really need to have defined methods of collaboration." (Undergraduate Bioinformatics Researcher, CA, 3/19/13)
- "…you need to get the idea that it is just a single entity; otherwise, the part of implementing a [computational] tool won't work at all." (Project Manager at EPCC, UK, 11/18/13)
- "You need a well understood objective. You'll be amazed by how many times, we don't know exactly what it is we're doing… So it sounds strange to say it, but a well defined clear objective… having clear objective is number one." (Computational & Domain Scientist, CA, 11/21/13)

Conclusion

In answering the research question, "*What key activities constitute productive collaborations among multidisciplinary participants in computational tool development and implementation in big data science?",* we concluded that it is important to appoint and maintain someone who knows how to communicate with both sides, create a common language that is used by all participants, and promoting a common goal for the team. These findings reinforce that "establishing mutual knowledge is important because it increases the likelihood that communication will be understood" (Cramton, 2001, p. 2). While these findings appear as common sense, many projects in practice often overlook these activities. Beyond the scientific computing context, these findings are applicable to projects that involve any two different fields of professional knowledge. A consultant would be able to use these findings to assist with a commercial and/or open source IT development project. Commercial companies can use these findings to assist with their behind the scenes development and implementation of technologies with lead users before rolling out a new product.

References

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative Sociology, 13(1), 3-21.
Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science*, *12*(3), 346-371.
Leonardi, P. M. (2009). Crossing the implementation line: The mutual constitution of technology and organizing across development and use activities. *Communication Theory*, *19*(3), 278-310.
Winner, L. (1993). Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, & Human Values,* 18(3), 362 – 378.