Building earth science cyberenvironments at the working collaboration level

Andrew R. Maffei, Heidi M. Sosik, Joseph Futrelle – Woods Hole Oceanographic Institution
Peter A. Fox – Rensselaer Polytechnic Institute

Background

Attempts to introduce new cyberinfrastructure in business and science environments often fail. The authors of the 2005 IEEE Spectrum article *Why Software Fails* [1] document a number of large and expensive business-related software development failures along with an analysis of the factors leading to those failures. The 2006 National Science Foundation Report "Cyberenvironment Project Management: Lessons Learned" [2] identifies a number of critical lessons on the topic of building science-related cyberenvironments:

1. Users and technologists need each other to succeed
2. You must have a target and know how to reach it
3. Leadership should be a partnership between technologists and domain specialists
4. Effective project management is essential at all levels
5. Communication is crucial
6. Good software development practices need to be established
7. Experiment-based software deployment is effective for helping users to own the software
8. Cyberinfrastructure is a living entity

The "elephant in the room" regarding the development of cyberenvironments is that establishing cultural shifts that enable interdisciplinary teams to work together to develop robust cyberinfrastructure is as important as, and perhaps more difficult than, simply building and deploying novel technologies. Bridging disciplinary cultures is essential to building cyberenvironments that meet scientists' real needs and naturally become a staple in their day-to-day use—even by those researchers generally not initially inclined to use advanced cyberinfrastructure.

The development of effective relationships and communication styles between information scientists and researchers using the cyberinfrastructure they build is vital to the success of such projects. Formally structured interdisciplinary design and development methodologies can help address these tough cultural and sociological challenges, leading to powerful new cyberinfrastructure tools and techniques along with an improved organizational ability to migrate discovered approaches and results from one science project for application in new areas.

Partnership forging with frameworks and methodologies

Substantial progress can be made on the lessons enumerated above by researching, testing, and applying frameworks and methodologies that help to build strong earth scientist - informaticist partnerships. Our experience is that such partnerships work. By including both skilled informaticists and a small number of earth scientists in projects that concentrate on solving the immediate, seemingly modest, informatics needs of individual earth scientists, (e.g., meeting basic data management and accessibility needs) the partnerships not only deliver improved science results early, but also catalyze broader and longer-term advances in the larger research community. Our evidence supports the importance of such initial efforts to build trust and strong working relationships; this investment can lead directly to information and activities that result in high impact cyberinfrastructures that are adopted quickly and readily by earth scientists.

In our pilot work in this area at the Woods Hole Oceanographic Institution (WHOI), we call these teams “Informatics Bridging Teams” (IBTs). IBTs provide a targeted context for informaticists and earth scientists to work to scope and develop pilot projects, establishing a foundation of trust and commitment that sets the stage for future collaboration on larger efforts.

We are in the early stages of working with the IBT concept, and experiences have been very positive. Our IBTs have already made significant progress in addressing the lessons enumerated above. Science users and the informaticists in the IBTs are very aware of their inter-dependency (lesson #1) because they are working closely together on the same objectives (lesson #2). These objectives are set through an iterative process that is fundamentally based on interactive dialog where science users articulate their needs, informaticists carefully listen and interpret those needs and prototype solutions, and the whole team evaluates progress before the next iteration (lesson #3). Effective communication and management is improved via an emphasis on skilled facilitation and a team ethos of full member inclusiveness in decision-making (lesson #4). In fact, in our experience, early investment in team building generates mutual respect such that team members actively seek and enjoy this iterative communication, and quickly appreciate that it produces successes more effectively and rapidly than other models (lesson #5). Within this IBT strategy, good software development practices are brought to the table by informaticists and progressively adopted by the science users as they see their value directly demonstrated in the informaticists’ work (lesson #6). Science users can adopt appropriate practices at a pace that is comfortable (one new practice at a time if needed) because the inherently experiment-based software development is evaluated and enhanced at each step (lesson #7).

Tools built by the IBT are customized for specific scientists’ needs, but informaticists’ involvement in use cases for other science laboratories means that shared requirements are easy to discover, and tools can be added to a common toolbox that eventually supports a variety of related science projects.

Working together with pre-existing cyberinfrastructure technologies and approaches to meet these more narrowly scoped science goals does not alone lead to the creation of a “transformative cyberinfrastructure”. However, a crucial outcome is that they produce the transformative foundation of trust, commitment, and mutual understanding necessary for success of these types of larger efforts. Also, somewhat unexpectedly, while informaticists and earth scientists are “in the trenches” working on smaller science projects, we have found that conversations emerge about larger cyberinfrastructure that would be helpful to the scientist and his/her broader community. As the IBT’s work unfolds, knowledge about informatics and earth science practices is transferred with increasing efficiency across the disciplinary boundary to enable faster and more effective development of generalized, shared infrastructure.

Thus, with careful attention to the framework and methodology adopted by such teams, it is possible to document use cases, requirements, and prototypes for a more sophisticated cyberinfrastructure. In our current IBT project, informaticists have been able to build elements of a larger community-scoped cyberinfrastructure while also meeting the immediate needs of the smaller number of scientists with whom s/he is working. A network of such informaticists could coordinate these activities and share tools and strategies across smaller projects.

Although it is likely that a variety of frameworks and methodologies can foster the kinds of partnerships that our IBTs represent, we have found that the Tetherless World Constellation (TWC) Methodology matches our organization's work style and culture.

Informatics Bridging Teams and the TWC Methodology

At the Woods Hole Oceanographic Institution (WHOI), we have recently begun (with support from the Gordon and Betty Moore Foundation) to develop and evaluate the IBT concept as applied to a set of related science projects. Our goal is to develop a strategy that produces demonstrable project-specific results in the short term, but that also lays the foundation for WHOI to more effectively integrate new informatics technologies into broader institution and community research programs in the future. Lacking our own computer science department, we have capitalized on a partnership with the Tetherless World Constellation at Renssalaer Polytechnic Institute (RPI) and have adopted their methodology [3] in our initial IBT activities.

The 11-component TWC Methodology is an iterative design and development methodology that focuses on the constitution of appropriate interdisciplinary teams with well-defined roles; the development and documentation of formal uses case based on explicitly-scoped goals and requirements; the development of formal information models and associated system models; and, the identification, development, and deployment of prototypes that serve as input into the continuing refinement of the use case and resulting system. The TWC methodology has been used in numerous scientific, technical, and other domains to build successful systems, including a number of earth science domains [3]. The IBT model that we are developing begins with training in the TWC methodology, followed by the development of formal use cases targeted at developing new systems to deliver improved science results.

Our first IBT is focusing on challenges in the area of ocean imaging informatics. The IBT includes a lead informaticist and three ocean scientists who previously worked independently on different aspects of ocean imaging. The three scientists have already led research efforts to build and operate different oceanographic instrument systems that collect large volumes of underwater imagery. These instrument systems are attracting interest from external researchers and resource managers who want to deploy them for a variety of applications and research programs. This has led to an immediate need to develop scalable informatics solutions that can accommodate many new users and ever-larger data sets. The instrument teams recognize that they face some common challenges and will benefit from a coordinated informatics effort to meet the demands associated with large data sets and computationally intensive analysis.

In the IBT context scientists and their staff provide domain expertise, and the informaticist and programmer provide informatics expertise. Other team members include a team leader/facilitator, a postdoc (yet to be hired) to work with scientists on new image processing algorithms, and an administrative staff person. Support is also available for software contractors with niche skills. A professor and staff members from TWC provide training in the TWC Methodology via workshops and consultation. Another group of students and a professor from the RPI computer vision laboratory add image processing expertise, providing improved code for calibration, color and illumination correction, segmentation, identification of regions of interest, and other image analytics requirements.

The project team aims to build an open-source, community-available, end-to-end infrastructure for collecting and processing oceanographic imagery, including data management, scientific workflow integration, provenance support, advanced image analytics, and support for data publishing and preservation. In addition it will improve software engineering by introducing revision control, issue tracking, automated testing, and continuous deployment. However, our first step towards these goals was not to start designing this all-encompassing cyberinfrastructure. Instead, we began by identifying lab-specific informatics challenges each of the scientists were dealing with on a day-to-day basis —their immediate "information headaches"—with the aim of addressing these smaller scoped issues as our first tasks.

Lab-specific challenges were identified during TWC-led training workshops that introduced the TWC Methodology to all IBT members. Scientists were encouraged to articulate and document real use cases with activity diagrams and information models. Participation of the team's lead informaticist was essential during the workshops. Using formal semantics, he was able to facilitate and direct the more advanced information modeling exercises with the ocean scientists. By reviewing the variety of use-cases from each of the labs, he was also able to identify the approaches and technologies appropriate for first iteration solutions to the more immediate informatics problems while considering their future applicability to larger scoped use-cases. By the end of the workshop, the scientists and informaticists worked together to select one use case—a web-based manual image annotation application—as a common need across the three labs that would be the IBT's first joint objective.

The ocean informaticist has been splitting his time between meeting the more immediate needs in each laboratory and working with the larger team on the design of the manual image annotation tool. By identifying and acting on "low hanging fruit" in one laboratory he has introduced significant performance enhancements to image processing code, and built web services that greatly simplify access to instrument data. He developed a prototype web-based visual summarization tool that presents web visitors with an overview of instrument imagery and metadata in near-realtime. Web services also provide machine-readable output, allowing instrument data to be syndicated to analysis services and other automated processing. These accomplishments have generated a high level of respect and interest from the ocean scientist. She is beginning to appreciate the value of informatics concepts and solutions she knew little about previously—global identifiers, web services, parallel processing, virtual machines with GPUs, etc. As an extra benefit, early success with applications in this laboratory has sparked even more interest and investment from members of the other two labs, who have now seen direct evidence that effective informatics solutions can accelerate the pace of research.

The development of the manual image annotation use case has enabled the IBT to begin work on a larger web-based system that includes manual annotation along with a variety of other capabilities applicable to ocean imaging instrument systems generally. Neither of these prototype tools were specifically identified as objectives before the IBT was formed; instead they have emerged as a result of the IBT's use of the TWC Methodology. Because the informaticist's work is coordinated across teams via the IBT, these "emergent software tools" will integrate with other solutions and contribute directly to the larger, community-scoped, goals.

Future work will involve investigating how the IBT framework can augment more traditional requirements-gathering methodologies. We believe that the experience of an informaticist working in an IBT can generate high-quality requirements for a larger cyberinfrastructure. Our experiences suggest that requirements can be generated more efficiently and completely in this manner when compared to the more conventional approach of bringing a large number of researchers together in nation-wide meetings to define requirements.

Interdisciplinary Networks for EarthCube

The EarthCube initiative provides an opportunity to build networks of informaticist / geoscientist partnerships, based on proven interdisciplinary frameworks and methodologies that can rapidly effect transformative change throughout the geosciences. WHOI's Informatics Bridging Team is one such framework designed to match a given organizational culture. It is likely that other organizational cultures may require somewhat different methodologies and approaches. EarthCube is a context in which informaticists can be trained in interdisciplinary approaches such as the TWC Methodology and can then be "embedded" in geo-science project teams, first to meet some basic information needs of geo-scientists, and then to use those advances to catalyze larger transformative efforts in the broader geo-science community.

As individual informaticists work in teams with geoscientists, they can also interact across these teams with other geoscience informaticists to discover common requirements and approaches. This would result in a library of real use cases and associated solutions that meet science and technical requirements. The documentation, approaches, experience, and tools produced in this process will provide critical and high-quality input into the design and development of larger, more broadly shared cyberinfrastructure. Effective sharing of both community-scale goals and project-scale tools across teams will ensure that many solutions can be combined or scaled up to meet larger scale needs.

We acknowledge close collaboration with Dr. Scott Gallager (BIO/WHOI), Dr. Hanu Singh (AOPE/WHOI), Prof. Charles V. Stewart (CS/RPI) and their laboratory staffs on this work.

References

[1] Charette, R., Why software fails. *IEEE Spectrum*, **42** 9 (2005), pp. 42–49.

[2] Spencer, B., Butler, R., Ricker, K., Marcusiu, D., Finholt, T., Foster, I. et al. (2006). Cyberenvironment project management: Lessons learned. Available: http://www.nsf.gov/od/oci/CPMLL.pdf. [Last referenced: 10.14.2011].

[3] Fox et al. (2011), An Open-World Iterative Methodology for the Development of Semantically-enabled Applications (in preparation).