GROUNDWORK
Furthering the Influence of Earth Science

We need to talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities

Matty Mookerjee*, Daniel Vieira, *Dept. of Geology, Sonoma State University, Rohnert Park, California 94928, USA;* ***Marjorie A. Chan,*** *Dept. of Geology & Geophysics, The University of Utah, Salt Lake City, Utah 84112, USA;* ***Yolanda Gil,*** *Information Sciences Institute, University of Southern California, Marina del Rey, California 90292, USA;* ***Charles Goodwin,*** *Dept. of Applied Linguistics, University of California at Los Angeles, Los Angeles, California 90095, USA;* ***Thomas F. Shipley,*** *Dept. of Psychology, Temple University, Philadelphia, Pennsylvania 19122, USA; and* ***Basil Tikoff,*** *Dept. of Geoscience, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA*

A UNIQUE OPPORTUNITY FOR CYBERINFRASTRUCTURE

It is increasingly important to integrate datasets and models from multiple geoscience subdisciplines in order to significantly advance our knowledge of how the planet works. To facilitate interdisciplinary investigations, geoscientists need a cyberinfrastructure that will easily access and combine datasets from all of the current and future geo-community databases. To this end, NSF introduced the EarthCube initiative (www.earthcube.org) to "create a community-driven data and knowledge management system that will allow for unprecedented data sharing across the geosciences." The ultimate goal of EarthCube is to transform Earth science investigations by promoting efficient data access, incorporating cyberinfrastructure into our scientific workflow, and allowing for increased sophistication of analyses and models (Gil et al., 2014; Kelbert, 2014; Richard et al., 2014). A significant strength of EarthCube is its potential to create sustained communication across the subfields within the Earth sciences, allowing scientists to ask new types of questions, and providing the means to address previously unanswerable ones. Examples of specific use cases are available on the EarthCube webpage; however, using machine learning to extract data from published articles (e.g., DeepDive [http://deepdive.stanford.edu]) and curating useful software/scripts (e.g., GeoSoft [http://www.isi.edu/ikcap/geosoft/]) are two widely applicable examples of EarthCube outcomes.

While the technical issues of interconnecting all existing community databases are significant challenges, an even more fundamental issue needs to be addressed: Not all communities have a database or the institutional support to manage one. In order for EarthCube to be successful, data from all subdisciplines need to be represented in the data management system. In particular, the field-based geosciences have lagged behind other subdisciplines with respect to developing a cyberinfrastructure for their datasets. This is likely due to the fact that most field data is collected and recorded in an analog format (e.g., Brunton compass, field notebook, and sketches) and through various personalized conventions. The additional step of digitizing these data is often onerous. In order to facilitate the development of cyberinfrastructure for the field-based geosciences, digitization processes must be incorporated into the typical geoscience workflow in a way that is as unobtrusive as possible (e.g., digital field notebooks, digital compasses, voice recognition software, digital pens, etc.). These potential solutions need to be developed in tandem with the cyberinfrastructure for managing these datasets. This is why it is critical to get the cyberinfrastructure and the field-based geoscience communities together and communicating effectively. Field-based geoscientists need the cyberinfrastructure community to advise them on the efficient collection of data for optimal digitization, while keeping them grounded in what is technically feasible. The geoscience community must engage and communicate their current and anticipated needs along with their specific data formats and requirements in order to design an effective data management system.

BRINGING CYBERINFRASTRUCTURE RESEARCHERS INTO THE FIELD

In order to facilitate the necessary communication between field-based geoscience and cyberinfrastructure communities, we proposed the currently funded NSF EarthCube project: "(EC3) Earth-centered communication for cyberinfrastructure: Challenges of field data collection, management and integration." With this project, we brought together various field-based geologists with computer scientists and a cognitive psychologist in a field setting. In August of 2014, 32 field-trip participants (12 computer scientists, 10 geoscientists, four graduate students, three undergraduates, two applied linguists, and one cognitive scientist) traveled to Yosemite and Owens Valley, California, USA, in order to discuss cyberinfrastructure-related issues. There is no better place to gain an appreciation for the field geologist's workflow than in the field itself. For the same reasons that we bring students into the field to explain fundamental concepts in the Earth sciences, the field provides an excellent venue for engaging with computer scientists about the multiple scales and interconnections of geological data, data collection strategies and techniques,

GSA Today, v. 25, no. 11, doi: 10.1130/GSATG248GW.1.

*E-mail: matty.mookerjee@sonoma.edu

and data representation. This field trip allowed these communities to understand each other's goals; geoscientists explained their data needs, while cyberinfrastructure experts explained the challenges associated with the current technology and data tools. Moreover, by presenting the results of data collected from the specific field areas that we visited, the computer scientists were able to gain an appreciation for the scope of scientific questions that geologists try to answer, the role of data in making geological inferences, the decision-making process for collecting future datasets, and how data and models interact in developing a regional conceptual framework. Sharing experiences in the field, both intellectual and practical, is a time-tested method for creating a strong and collaborative scientific community.

SURVEY OF FIELD-TRIP PARTICIPANTS

Approximately three weeks prior to attending the 2014 EC3 field trip, the participants were asked to complete a voluntary, 55-question survey in order to assess their current understanding, interest, and experience with cyberinfrastructure in the geosciences as well as their knowledge and level of involvement in EarthCube. Of the 32 field trip participants, we received 25 responses to the pre-field trip survey. Directly after the completion of the field trip, the participants were again asked to complete the same survey along with an additional 15 questions. The post-field trip survey garnered 28 responses. The goal of these two surveys was to evaluate the effectiveness of using field experiences to help identify and bridge the conceptual gaps between the geoscience and cyberinfrastructure communities with respect to developing the appropriate data infrastructure for the field-based geosciences.

We highlight here the most salient results from these surveys, summarized in Figure 1. The complete survey and partially redacted results can be downloaded from this link: http://www.sonoma.edu/users/m/mookerje/EC3_pre&post_survey_Redacted.pdf. The data suggest that after the field trip, both geologists and computer scientists felt that they had a better understanding about the current challenges facing field geoscientists with respect to getting their data fully incorporated into the appropriate data infrastructure. Similarly, both groups also had a better understanding of what was needed by the cyberinfrastructure community in order to design an effective integrated data system. Computer scientists gained a more complete idea of how they could help geoscientists capture their data into a data management system, and they were confident that it was possible to fully capture the types of data used by field geologists into an integrated data system. It is unsurprising that this increase correlates with an increase in the computer scientists' understanding of what geologists do in the field. After the field trip, both the geoscientists and the cyberinfrastructure participants responded that they were more likely to utilize EarthCube resources in their own research and that they believed that the development of cyberinfrastructure would ultimately save them time in their various scientific endeavors.

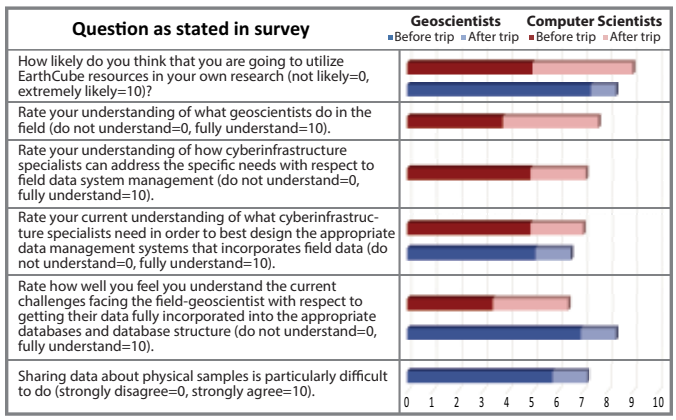

Figure 1. 2014 EC3 Field Trip Pre- and Post-Survey Results.

THE PATH FORWARD

For EarthCube to succeed, it needs to engage the larger geoscience population and build consensus around what cyberinfrastructure is the most appropriate for our community. Activities designed to communicate across disciplinary lines—such as the field trips described here—are an effective approach. However, we also need mechanisms that scale in such a way as to incorporate input from a larger percentage of our scientific community. From our perspective, almost every geologist that we talk with recognizes the need for cyberinfrastructure. It is clear that we must have digital data systems in order to increase the capacity and efficiency for conducting science, improving the quality of science, and facilitating new discoveries. Further, there is increasing pressure from funding agencies that we make our data more widely available. Regardless of this pressing need, many of the geoscientists we have talked to seem to be taking a "wait and see" attitude toward cyberinfrastructure and EarthCube specifically. While that reluctance is, to some extent, understandable, cyberinfrastructure cannot be designed well without input from the communities that it is meant to serve. We are now at a critical junction to make field-based geologic data an integral part of EarthCube. This task will be difficult and will take time away from other activities, but it is essential to the future of the science. We urge all geoscientists—especially early-career investigators and graduate students, who will inevitably take advantage of these systems during their careers—to engage and help shape the emerging geological cyberinfrastructure (http://earthcube.org/info/get-involved). One excellent venue for learning more about ongoing and planning activities is at EarthCube Town Hall meetings hosted at the national GSA and AGU conferences. However one decides to engage, now is the time for geoscientists to make their opinions, perspectives, and data needs known.

ACKNOWLEDGMENTS

M. Mookerjee's work is funded by NSF EarthCube grant ICER-1340265.

REFERENCES CITED

Gil, Y., Chan, M., Gomez, B., and Caron, B., editors, 2014, EarthCube: Past, Present, and Future: EarthCube Project Report EC-2014-3, Dec. 2014, http://earthcube.org/document/2014/earthcube-past-present-future (last accessed 25 June 2015).

Kelbert, A., 2014, Science and Cyberinfrastructure: The Chicken and Egg Problem: Eos, v. 95, no. 49, p. 458–459, doi: 10.1002/2014EO490006.

Richard, S.M., Pearthree, G., Aufdenkkampe, A.K., Cutcher-Gershenfeld, J., Daniels, M., Gomez, B., Kinkade, D., and Percivall, G., 2014, Community-Developed Geoscience Cyberinfrastructure: Eos, v. 95, no. 20, p. 165–166.

Manuscript received 27 Feb. 2015; accepted 9 June 2015. ✸