Opportunities and Challenges in IIS-Enabled Scientific Discovery

Liz Bradley

University of Colorado

lizb@colorado.edu

In my opinion, there are significant strategic and cultural challenges involved in figuring out how IIS research could contribute to scientific discovery—challenges that must be part of the conversation from the beginning. There are a tremendous number of science discovery opportunities out there, and a whole host of good computer science approaches to those problems, dating back to BACON *et al.* and extending to the powerful and interesting ideas that are in the statements of the people at this workshop. In view of that history and potential, it goes almost without saying that IIS can help with structuring problems, creating models, managing hypotheses, designing experiments, configuring instruments to collect the right data, understanding data and analytic processes, and so on. The challenge will be to pose, focus, and operationalize the effort within the existing culture in (and between) the research community and the funding agencies.

The key to launching a successful IIS/discovery research program will be not only to formulate and write a solicitation with a solid, coherent, compelling message, but also to set things up so that the research community and the funding agencies can follow through in a manner that is truly consistent—in practice, in science, and in philosophy—with that message: i.e., in a manner that doesn't regress to the mean of "the way it's always been done." IIS applications in scientific discovery are a wonderful, rich, important research area. But they have the classic challenge of interdisciplinary research: they risk being perceived as "just" applications of CS to problems in other fields. If an IIS/discovery initiative is to work, the PMs and the PIs must have radically different expectations of the proposal process. The PIs will need to stick their necks out and propose the kinds of truly transdisciplinary ideas and approaches that make real scientific discovery possible; the PMs will need to have the autonomy—and the mandate—to make funding decisions that are in line with that mission.

I've done a lot of interdisciplinary work on scientific discovery, so I have some experience with these challenges. This dates back to the early 1990s and the PRET system (e.g., [3]), which discovers differential-equation models of black-box systems. PRET was funded by young-investigator type awards that granted me the kind of interdisciplinary freedom that is necessary to do this kind of work. More recently, I worked with a software engineer and a bunch of geochemists to develop a software tool called ACE (e.g., [4]) which reasons about paleolandform dating. ACE takes as inputs the exposure ages of a set of rock samples taken from a landscape, derived via cosmogenic isotope techniques. It answers the scientific question "What geological processes could have produced this situation, and over what time scales?" ACE remains in active use to this day, several years after the development cycle ended, without a single request for help from the geochemists to the

computer science side of the team; its website [1] has received over 11,000 hits in the last year, and the software (20,000 lines of python code) has been downloaded 270 times—a significant number in a research community of $O(10^2)$ PI-level scientists [2]. That project was funded by an ITR, another novel and explicitly interdisciplinary initiative. Lately, I've been working with a team of geoscientists to put together a proposal about scientific discovery in the context of paleoclimatology. This project, whose goal is to build age models for core data (ice, ocean sediments, etc.), would remove the main current roadblock in paleoclimatology. You need an age model—a curve that relates the depth in the core to the age of the material at that point—to do science on a core, but building those models takes a fair bit of time and requires expert knowledge. Because of this, there are vast archives of raw paleoclimate data lying around waiting for analysis. The World Data Center for Paleoclimatology archive at NOAA, for instance, contains millimeter-by-millimeter measurements of up to 13 variables in cores from 7000 sites, some of which are thousands of meters in length. Without intelligent computational assistance, needless to say, this is not a humanly possible task.

Paleoscience is only one of many instances where there is significant opportunity—and *need*—for IIS to contribute to scientific discovery. But we will have to make it very clear what the goal of this initiative is. To aid in scientific discovery? Or to develop new computer-science techniques that are motivated by or could be used for scientific discovery? These will produce different results, and you can't do both. 110% only exists in sports journalism. A project team can't put 100% effort behind *both* the application *and* developing new CS. If we want to foster advances in IIS techniques for scientific discovery, we will have to make the point, the philosophy, and the goals very very clear—and then somehow make sure that the research community (and the review process) operationalizes them faithfully.

And that may well be the most difficult (and radical) part of all of this.

References

[1] `ace.hwr.arizona.edu`.

[2] `cosmognosis.wordpress.com`.

[3] E. Bradley, M. Easley, and R. Stolle. Reasoning about nonlinear system identification. *Artificial Intelligence*, 133:139–188, December 2001.

[4] L. Rassbach, K. Anderson, and E. Bradley. Providing decision support for cosmogenic isotope dating. In *Proceedings of the 22nd Conference on Innovative Appliations of Artificial Intelligence (IAAI)*, 2010.