

## NIH Conference on Knowledge Environments for Biomedical Research (December 11–12, 2006)

Partha P Mitra

Published online: 9 May 2007  
© Humana Press Inc. 2007

When writing the poem “The Rock,” T. S. Eliot probably didn’t realize that a phrase from it would end up as an internet cliché (“Where is the knowledge we have lost in information?”)—a web search with this exact phrase returned 36 thousand hits. The phrase, eloquent even after overuse, set the stage for an NIH conference on Knowledge Environments for Biomedical Research, held December 11–12, 2006. About two hundred and fifty attendees, virtually the who’s who list of the subject area, drawn from academia, industry, and government attended this 2-day conference in Bethesda, MD.

The goal of the conference, organized by staff members from the NIH, was to “elucidate strategies to develop, sustain, and enhance knowledge environments for biomedical research.” The first day of the conference saw a series of talks, followed by breakout sessions on the second day where attendees brainstormed about such strategies, culminating in a series of short presentations and a panel discussion. This news item represents an independent opinion from the perspective of an individual attendee. An official report of the results of the meeting will be made available by the organizers at a later date. Materials from the workshop are available from the conference web site (<http://www.esi-bethesda.com/ncrrworkshops/kebr/>).

What is a knowledge environment? This was an underlying, recurrent question that was asked and answered multiple times during the 2 days from different perspectives. The answers ranged from “you know it when you see it” to more elaborate characterizations. One common theme

was that a knowledge environment is more than a simple collection of data sets and embodies some understanding about the domain in question. A field or scientific community may not be ready to have a knowledge environment unless there are common “fiducial” concepts, such as genes. This is a vexing question; in neuroscience, for example, such shared conceptual frameworks may or may not yet exist for the more fragmented branches of the discipline. A knowledge environment collects together digital resources and tools (databases, controlled vocabularies, analysis tools, user interfaces, and so on) based on such common concepts or understanding. Definitions aside, there was general agreement that knowledge environments continue to be vital to biomedical research, and much of the workshop was devoted to strategic “how” questions.

The first day began with a talk from Michael Huerta (Associate Director for Scientific Technology Research, NIMH) defining the charge to the workshop. Dr. Huerta emphasized that top-down, mandated approaches and bottom-up, purely community-driven approaches both have their limitations, and that an intermediate approach with guidance and support from above, and leadership and participation from below, may be the best road forward, although the success of this middle road would depend on dedicated players.

After the introductory talk, there were three sessions, each with a plenary talk followed by panel remarks and discussions with the audience. In the first session on information representation, James Brinkley from the University of Washington spoke about anatomical structure as a conceptual framework for integrating heterogeneous brain imaging and electrophysiological data. The second session on user-driven design saw a presentation by Ary Goldberger from Harvard Medical School on Physionet and related tools. The third session on sociocultural considerations included a presentation by Lawrence Hunter from the

---

P. P. Mitra (✉)  
Cold Spring Harbor Laboratory,  
Cold Spring Harbor, NY 11724, USA  
e-mail: mitra@cshl.edu

University of Colorado Health Sciences Center, in which he stressed the need for fiducial concepts as prerequisites for knowledge environments. Gary Olson from the University of Michigan sounded a cautionary note about the success rates of “collaboratories”: in brief, most such efforts in the past have failed.

The first half of the second day of the workshop was spent in brainstorming sessions in breakout groups. The charge given to the groups was to help elucidate strategies to develop, sustain, and enhance knowledge environments for biomedical research. Each group tackled the same set of six questions in differing orders. How do you decide when a community is ready to benefit from a knowledge environment? The answer clearly depends on the level of commitment: a no-frills repository has a lower bar than a persistent knowledge base with high expertise curation, so the decision criteria will differ. The field has to be sufficiently mature and interactive, as indicated by the existence of a shared conceptual framework or vocabulary, or even better, one or more promising knowledge environments already in development. On the other hand, recognized nomenclature confusion and the need for pedagogical tools for newcomers to the field may also indicate that such efforts will be fruitful.

Who should be involved in the development? A combination of resources, expertise, and community need are important determinants, and coordination between the different actors is necessary. One message came across loud and clear throughout the meeting: scientific curators are essential to the enterprise, and better support and reward mechanisms are needed for this relatively new scientific “caste.” How and when do you solicit community input? It was pointed out that biologist end users and librarians were groups not well represented at the meeting, and the need for more outreach and market research was emphasized. Efforts could include smaller focus groups and other active mechanisms at the NIH to solicit community input, and perhaps also the formalization of explicit marketing and outreach plans in grants for knowledge environment development.

What are the criteria for success and signs of failure? Different mechanisms are needed to evaluate infrastructure development and user-oriented tools. Some quantitative metric is clearly needed to evaluate usage and impact, although a simple counting of downloads or hits may not be enough: there is need for the development and promotion of a standardized citation mechanism for knowledge environments and similar resources. Silence, in the form of lack of publications, low usage statistics, and so on, is a sign of failure, unless such a fallow period is explicitly anticipated during initial developmental stages.

How can knowledge environments coevolve? Can or should they eventually merge? While opinions differed,

there was an overall sense that making interoperability simple is a better goal than merging into one large knowledge base. In this context it would be important to study some models, such as standards in industry and in governmental agencies, as well as Internet success stories such as Wikipedia or Google that have reshaped the landscapes in which future knowledge bases will live.

Finally, how can information about knowledge environments be disseminated? There was agreement for a regular, predictable venue where a catalogue of resources can be found; maintaining a central index or database of this sort would perhaps be the most effective way to ensure dissemination. Scientific curators who speak the language of the target audience are also important in this regard. A low-barrier digital library at the NIH could serve as a default, persistent data and software repository, and perhaps the NIH itself needs to play a more active role in such dissemination through its program staff or through organizing meetings.

Some themes came up multiple times during the question-and-answer periods. For many practicing biologists, spreadsheet software has often served the basic functionality that a knowledge base ultimately provides. At what point does it become worthwhile to migrate to more sophisticated platforms, and how does the biomedical researcher acquire the relevant expertise to use more sophisticated tools? In general, the need and scope for mutual education between the biomedical researcher and the software engineers involved in knowledge engineering was repeatedly emphasized. Scientific curators play a crucial role in providing this interface. It was also noted that the scientific literature is still, in some sense, the best and most comprehensive knowledge environment available to the community.

Another issue inspiring debate was that of how best to achieve standards in data formats and controlled vocabularies: should these be mandated, or allowed to self-organize from the community? What is the correct middle ground? As with many questions considered at this conference, the answers will vary depending on the maturity of a particular subfield of biomedical research. However, during these 2 days, one commonality was very clear: that all the participants in the conference had gained greatly from the discussions and debates and from exposure to the broad spectrum of perspectives represented. The attempts by the conference organizers to provide order in this dynamic and complex research area are worthy of applause. Possibly the best measure of success of the conference was the feeling among several participants that the group should reconvene in the not-too-distant future. After all, the best knowledge environment of all is a group of experts in collegiate conversation.