NET*Working*

'Incubation' for a New Idea

Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations

Karen Baker (UCSD, Palmer LTER) and Geoffrey Bowker (UCSD, Communications Department)

As more scientists exchange and post data digitally, the challenges of transforming data systems into knowledge systems come into sharper focus. As communities address issues of data availability, the understanding of data incompatibility issues deepens. The LTER integrative approach to science brings with it a familiarity with difficulties arising from technology with differences in hardware, software, and formats as well as from uniformity or standards with differences in collection units, classification categories, and methods.

There are a variety of challenges to data integration arising from creation of appropriate metadata, the descriptive documentation about the data. The 'context' of data is complex: one aspect of a dataset's metadata relates to the environment it describes while another aspect relates to its creator and is impacted by its association with a site, project, repository and network. As a result, in addition to the barriers of technology and uniformity, we identify within the multiple layers of metadata that there are barriers of social organization to consider.

A recent NSF report (Kinzig et al. 2001) focusing on priorities for interdisciplinary environmental research presents the human interface with the environment representing humans as agents within the environment as well as in its perception. The report calls for careful consideration of how scientists negotiate their data. How the environment or the data is reported with its attendant metadata impacts ultimately how it is perceived and reused. The realization that metadata are not socially or culturally neutral improves our ability to design knowledge systems. Techniques used in the structuring of information include classification schemas as well as attention to semantic vocabularies and domain maps to describe the context of the data.

We ask whether there are methods complementary to logic-based approaches to information retrieval that can include data collection practices, project summaries and research vignettes.

We were recently awarded a Biodiversity and Ecosystem Informatics (BDEI) cross-agency (NSF/USGS/NASA) (Maier et al. 2001) incubation grant to consider impacts on metadata and infrastructure of the heterogeneity of 1) data, 2) collaborators, and; 3) organizations. This project seeks new ways of grounding environmental data with metadata so that data can be used more flexibly today and also be available over the long-term in a form usable for unanticipated future queries. The goal of this project is twofold: to open up a new field of database enquiry tied to the specific challenges of biodiversity and ecosystem science and to initiate a cross-domain dialogue between LTER ecologists (Callahan 1984), information managers (Baker et al. 2000), and social scientists (Bowker 2001).

This BDEI incubation grant will support a postdoctoral fellow trained in participatory design techniques who will focus on a prototype project as well as a community forum. Work begins in January 2002 with arrival of Dr. Helena Karasti (2001) who recently finished her thesis work in the Department of Information Processing Science at Oulu University in Finland. With experience in collaborative design, she will work with selected data from the Palmer LTER while considering the data capture and metadata roles in transforming data into knowledge. This project will develop into a larger follow-on study of modes of packaging data. This work facilitates a timely dialogue focused on "data ecology" (the relationship between data and their multiple environments) and builds toward the concept of an "organizational ecology" (the relationship between data and participants and their networks). The potential impacts are high in terms of both network infrastructure and expert systems.

The LTER Network fosters an environment that engages research scientists in the process of preserving their data digitally for the long-term. The vision broadens and deepens with collaborative participation in metadata and knowledge system design. An active partnership spanning several communities (NCEAS/SDSC/LTER) currently is exploring a metadata-based framework and includes participation by the LTER Central Arizona-Phoenix site exploring structure and developing tools. Our BDEI project joins these informatics efforts. Given the many facets of human knowledge, it will take a team bridging data-creators and data-managers, working in collaboration with data-handlers as well as those familiar with data-perception and data-context to address data integration in the short-term and data reuse in the future.

The LTER Network is well suited to initiate studies on the use of technology and standards as well as on the impact of social organization on knowledge systems.

References

Baker, KS, BJ Benson, DL Henshaw, D Blodgett, JH Porter, and SG Stafford, 2000. Evolution of a multisite network information system: the LTER information management paradigm. BioScience 50(11): 963-978.

Bowker, GC, 2001. Biodiversity, Datadiversity. Social Studies of Science, 30(5), 643-684.

Callahan, JT, 1984. Long-Term Ecological Research. BioScience 34(6), 363-367.

Karasti, H, 2000. Increasing sensitivity towards everyday work practice in system design. Unpublished PhD thesis. http://www.tol.oulu.fi/~helena/

Kinzig, AP, SCarpenter, MDove, GHeal, SLevin, JLubchenco, SHSchneider, DStarrett, 2000. Nature and society: an imperative for integrated environmental research Excecutive Summary of a workshop sponsored by NSF, Developing a Research Agenda for Linking Biogeophysical and Socioeconomic Systems, Tempe, AZ, 5-8 June, http://lsweb.la.asu.edu/akinzig/report.htm

Maier, D, E.Landis, JCushing, A.Frondorf, ASIberscatz, MFrame, and JLSchnase, 2001. Research Directions in Biodiversity and Ecosystem Informatics, pp30. Workshop held at NASA Goddard Space Flight Center, June 22-23, 2000, Greenbelt, Maryland.