Biodiversity and Ecosystem Informatics (BDEI-3) Workshop Summary
Eco-Informatics and Decision Making
The Evergreen State College - Olympia, Washington
December 13-15, 2004
http://www.evergreen.edu/bdei

Decision makers at non-governmental organizations and at all levels of government often face significant information technology (IT) problems when integrating ecological or environmental information to make decisions[1]. This workshop aimed to identify new computer science and social science research and technology that would support these decision makers and their information providers. The resulting report will be used by sponsoring agencies (and others) to guide future IT research and development. In particular, the NSF plans a Request for Grant Proposals, and subsequent awards, in 2005.

This workshop was funded by NSF (NSF IIS 0505790) and the USGS/NBII, with significant input from NASA, the EPA, USFS, USPS, and NGO's (The Nature Conservancy and NaturServe).

Day One Evening - Setting the Stage. This after-dinner kickoff plenary session set the context, goals and objectives for this 1.5 day workshop.

- BDEI-3 Goal: Define a computer science and social research agenda to ameliorate IT problems faced by decision makers who use eco-informatics artifacts.

- The BDEI-3 report will serve as basis for future NSF R&D that will form the basis of a research and policy agenda in biodiversity and ecosystem informatics. It is expected that other agencies will also earmark funding for this area.

- BDEI-3 complements BDEI workshop reports from previous years, but is significantly different in that previous workshops did not address IT needs of decision makers.

    - The BDEI-1 workshop and report (2001) set forth a national BDEI research agenda to address infrastructure challenges, next-generation CS/IT applications, data robustness. It also reiterated the need for government and industrial support of both CS/IT and environmental research. This workshop led NSF to fund 15 BDEI planning grants.
    - In 2003, the NSF BDEI PI's meeting (dubbed BDEI-2) reported BDEI-1 grant results to program directors at NSF, USGS, and NASA. These findings set the basis for future CS/IT research prospects.

- The 1998 PCAST Report characterized bioinformatics as a biology and CS/IT cross-discipline, recognized the biodiversity-ecosystem nexus as an information enterprise, and envisioned analytical and synthetic capabilities among other foci in the next generation of NBII-2 information services.

- Three NSF Digital Government research case studies (DoE, EPA-Air Networked Environmental Information System for Global Emissions Inventories, and eRulemaking) illustrated the social science importance of problem selection, agency contact, collaboration, mutual expectations, and proposal strategy as factors in cultivating a successful digital government research experience (Univ. S. Calif.).

- EPA science advisors, in an effort to gauge pesticide exposure risk, have given preliminary consideration to CS/IT software that has the potential of avoiding protracted disputes regarding assumptions about public health risks.

[1] By "decision makers" we mean those in national, state, and local government, and NGO resource managers, who manage natural resources or carry out policy, or provide policy information to those who do.

Day Two – Defining The Problem Space

- A Plenary Panel focused on the key players in eco-informatics decision making and how data are used and provided to government and NGO decision makers:

  - Policy makers and their clients' information needs; communities of interest as well as of place; equitable information access; mass customization; distinguishing measurements, indicators, and interpretations; and metadata and validation are all essential to BDEI-decision making. "Better data lead to better dialogue, which leads to better decisions." (USFS)

  - State/EPA Environmental Information Exchange Network improves secure data exchange and timeliness between states and the EPA via web services, and facilitates adoption of new standards. Fish Tissue Contamination and Birth Defects Assessment is a key application. (ECOS)

  - Conservation Informatics is hard, and Data and Tools form a demand cycle. Biodiversity data management and collection would be more efficient were data managed in common formats, with better decision support tools, e.g., a common framework for geographic information. (NatureServe)

  - NASA's Science Mission's experience with decision support for earth science, in particular the invasive species project, shows how remotely sensed raw data (observations) can currently be used in conjunction with models (predictions) as input to decision support tools. (NASA)

  - An NSF Digital Government research case study (Coastal Management) showed that integrating multi-source spatial information requires enhanced data handling capacity, coastal geospatial information for intergovernmental agency operations, and innovative modeling. Detailed geospatial information and accurate modeling provide considerable potential for monitoring and prediction of coastal ecological changes. (Ohio State)

  - An NSF Digital Government research case study (Oregon Coastal Atlas) relies on resource decisions at local levels, where decisions are fraught with problems in assuring accurate resource status information. Collaboration through improved computational infrastructure and community partnerships has yielded success with hazards management, watershed assessment, and ocean protection programs. More research needed for improved search. (Oregon State)

- Breakout Session I – Five groups met independently to articulate real-world problems faced by providers and users of ecological and environmental information:

  - Policy problems that organizations (across all sectors) encounter in eco-informatics include: provision (e.g., funding), production and maintenance of eco-informatics tools and information; use and abuse of eco-informatics tools and data; cross-organizational sharing of eco-informatics data and tools, and communication of analysis-grounded management decisions.

  - Data Presentation problems arise from complex interactions between user needs (nature of required task and time involved) and data (i.e., metadata, raw data, accuracy specifications, methods, documentation, policy). System limitations (e.g., software modalities, availability and costs of hardware); and information format further complicate presentation. Critical research includes determining what information is best on which medium, cross-referencing and supporting data across presentations, representing time and change, new media (e..g, 3D, VR), and user task definitions.

- Geographic Data Gaps between biodiversity-rich and conservation-managed land areas impact decision making. Problems stem from: lack of needed data sets or access to them; disjunct data sets that require manipulation to compensate for temporal and/or spatial gaps); emphasis on adaptive management which out-paces data reliability; and lack of an adequate professional experts database.

- Tool problems involve: lack of a tool "clearing-house"; use with different data or new data types; data collection (e.g., GRID computing); lack of user frameworks and product suites; need for better tools to support metadata issues (creation, quality, etc.); need for improved standards in development; social science issues of usage, sharing, and adoption.

- Indicator problems exist because: indicator definition, relevance, and value are neither well-defined nor communicated; constituents may be uneasy with environmental measures; data gaps effect reliability and trust. Inherent complexity of the ecosystem/human health dynamic further complicates this issue. Examples include: Death Valley Pupfish or Washington Shellfish Bed Closures.

Day Two (cont'd) – Articulating the Research Issues

- This plenary session focused on how current research projects employ interdisciplinary approaches involving government partners to solve problems similar to those identified in the first Breakout Session:

  - An NSF Digital Government research case study (Forest Portal) showcased an adaptive management tool that harvests information to sustain forests; this highlighted importance between Federal agencies and academic institutions, and demonstrated the capabilities of using metadata attachments. (Portland State)

  - An NSF Digital Government research case study (UrbanSim) highlighted how ecological models and establishing partnerships contribute to data collection, preparation, and assessment. These factors will likely lead to realistic policy scenarios and major policy applications in 2005. (U. Washington).

  - An NSF Digital Government research case study (Understanding Government Statistics) aims to model user access to U.S. government statistical information to integrate data across agencies. In building a system that harvests government web pages, challenges were to find data that map to user requirements and design an interface that relies on metadata generated from the web sites. The value of working prototypes was emphasized.

- Breakout Session II – Focused on the relationships of the following five information technology areas to eco-informatics problems identified in the morning Breakout session, thus defining BDEI-3 research issues:

  - Social and Human Aspects of Eco-Informatics and Decision-Making aka *Human Centeredness* – research issues included: Collaboration in eco-informatics tool development and information sharing among decision makers (e.g., measuring success, determining appropriate institutional designs and incentives or disincentives); human-computer interaction (human/tools interface); impact on management practices; education and training (data management domain procedures); and user needs (user requirements, system design).

  - Modeling/Simulation research issues included: Coupling diverse models; addressing values in design (models for diverse stakeholders); incorporating new visualizations for model results; representing error and uncertainty when presenting information to decision makers; challenges in handling large data sets; and open source modeling infrastructure.

  - Data Quality research issues were summarized as how to determine and communicate uncertainty to decision-makers when they use multiple data sources. Methods are needed to mitigate introducing error

when creating and combining data sets, and to associate error with alternative decisions. The question of whether metadata could become an obligatory part of the data set was raised.

- Information Integration involves mechanisms for reliable, transparent and authoritative data combination. Associated research issues include: Defining the dimensions of integration; quantifying semantic distance; integrating multiple ontologies; promoting document modeling; evaluating utility of qualitative and quantitative data; need for tools to support data integration; and how one evaluates knowledge from non-traditional sources.

- Ontologies are useful in providing metadata (semantics) over databases, making cross-disciplinary connections, and thesauri. Ontologies on the Grid would help users find data and functionality. Tools to build, verify and deliver ontologies still require considerable research. Other phenomena that require research are understanding gaps and inconsistencies in ontologies, trust and verification of the content of ontologies, and understanding and handling change in the material represented by onotologies in ways that go beyond simple versioning.

Day Three – Refining the Research Agenda

- Senior scientists and decision makers from a variety of disciplines commented on and critiqued the breakout group reports from the previous evening:

  - Communication enables collaboration (human centeredness), trustworthiness (ontologies), and data sharing (data integration). Social science is characterized by indigenous local/community knowledge + ethics of decision making (data integration), and user needs (the futures market). Ontologies, coupling diverse models and how second and third generation metadata can be used to define data quality are particularly important. (Bowker)

  - A fictitious, ideal decision making tool *Yoda* helped define decision makers as those who choose among the alternatives, and what they do as integrating – via sharable data structures, compatible software tools, human collaboration, and understanding outcomes. This is an awesome task that involves ontologies, semantic distances, data quality assessment, etc., and many complex steps. (Tosta)

  - The sheer number, breadth and complexity of problems and potential solutions suggested at this workshop dictate that it will take decades to solve them – all while species and ecosystems disappear at an increasing rate. Thus, we need to prioritize the critical informatics problems – ask where do these problems intersect across agencies and environments to find the greatest synergies, ask which of those with the greatest intellectual merit could be solved with focused R&D, and ask where public and private funds could be leveraged? A follow-on workshop of eco-informatics professionals and computer scientists, itself followed by an online survey, auction or futures market could define this. Because problems are both technical and sociological, a few well chosen broad projects in those areas could serve other more focused research. (Biasi)

  - One senior scientist brought up the issue of feedback loops, and urged that if a resource manager becomes more effective at what they do, then the effect of the manager on the system he or she manages is not negligible. We know very little about this problem. (Rossignol)

- European views on eco-informatics and decision making conveyed during the meeting noted that eco-informatics are treated as general beneficiaries of the National funding programs within several EU Member States; EU Priorities within eco-informatics suggests strong policy association with problem solving; EEU and USEPA cooperate on several general topics including environment and health – related

issues and indicators. The Dg.o 2005 annual conference will host a workshop to promote EU/US eco-informatics research.

- Breakout Session III – Breakout groups from Session II reconvened to identify a real world problem where the research they proposed would be particularly helpful, or to articulate research infrastructure needed to sustain that research: [I re-ordered these to match the order of Breakout Sessions II –jbc]

    - Social and Human Aspects of Eco-Informatics and Decision-Making aka. *Human Centerdness* – Advancing the eco-informatics agenda hinges on both new technologies, and new understandings of how information infrastructures inter-relate between individuals, organizations, communities, disciplines, information resources, and tools. Consider State Agency Official "Jane Doe" prioritizing parcels for conservation. She is interested in forecasting land use change over a region with the hope that that will identify habitat parcels most threatened by human encroachment. Ideally, Jane would like some kind of policy-relevant modeling capability to help identify the "development fringe", but she cannot develop that on her own. Because others, whom Jane might not even know about, may be well on their way to doing this, tools to facilitate the investigation would include library management systems, and newer, innovative collaboration tools and computer-based land use change models. The breakout group considered this scenario is as it would play out now, and in ten years if the recommended research were successful. A second scenario involving the Death Valley National Park Devils Hole Pupfish illustrated an immediate need for tools that integrate information over time and across agencies, evaluate legacy data, identify indicators, visualize alternative actions, model current ecological conditions, and find similar studies.

    - Modeling/Simulation – This group emphasized research issues in simulation and modeling (model coupling, the issue of values in design, improved visualizations for model results, the ability to represent error and uncertainty, to handle large data sets). They proposed an open-source, flexible, reusable modeling infrastructure, along with the social practices that sustain it. This would allow researchers and decision makers to experiment freely with new models and/or change existing ones.

    - Data Quality – Given that the problem of data quality in decision making is how to determine and communicate uncertainty to decision, the research question is whether uncertainty associated with data quality and synthesis really influence policy and planning? There are two issues: where diverse data sources are combined and how metadata can become a part of the data set. Stories corresponding to the research issues were articulated. Why couldn't NSF publish metadata standards across all grants, instead of just certain programs? The research challenge is how general can these tools become? Could they be applied to a wide range of ecological datasets?

    - Ontologies/Integration – Story: The reclassification of rainbow trout as salmon in the early 1990's and a subsequently implemented information system,had broad-reaching effects; moral: no indicator is innocent, and IT systems have social consequences. Data collection, ontologization, and modeling embody value judgments; how can computer scientists and developers be sensitized? Semantics of BDE is critical – defining and operationalizing meanings, data standardization, semantic services. Transfering knowledge from other domains to BDE is itself research. Quality control, data access, and collaborative decision making support are also critical. Future IT applications should warn scientists and policy makers of impending circumstances.

Major Conclusions and OverArching Themes:

1. Decision Making. In addition to the predetermined major research themes, those conducting research in this area should have an understanding of decision making in general, specifically in the context of eco-informatics. A task force was assigned to report back to workshop participants, on the nature of decision making as "approaches to help actors make decisions among alternatives". The team noted that the domain is particularly difficult because of its nature of including both public policy and the complexity of eco-systems. One point of contention is that environmental situations issues are complex and there are considerable uncertainties, but in political and policy situations many decisions are placed in a "yes or no" context. (Fülöp, Roth, Schweik, Bowker).

2. Unique Nature of BDEI & Decision Making Research: While much of the information technology and social science research needed to solve the area's IT problems is the same research blocking progress in other domains, there are some special issues that arise here. Work in this area must take into account the need for combining quantitative information with qualitative, the social and economic value associated with these decisions, and the risk involved in using information technology to make resource management and environmental decisions that could potentially significantly impact health or must stand up in court.

3. Graduate Student Report. Workshop logistics were carried out by graduate students from Evergreen Masters' Program in Environmental Studies. These students also attended workshop sessions and reported that involving students in this research and using their research as a teaching tool requires trans-disciplinary communication, the development of new methods for collaboration, defining an integrated language, information dissemination and the development of eco-informatics educational materials. Ethical issues around large data repositories was seen as a particularly fruitful area. Approaches for involving students were suggested, and funding interdisciplinary mission-oriented task forces to address local problems was seen as a way to pursue these goals. The students encouraged NSF to partner with agencies that support applied student research.

4. Sustaining the Research: Funding agencies must work together and with principal investigators, information managers and decision makers in **sustaining and encouraging innovation, research and development** in this area. Considerable attention should be paid to assuring a cycle of innovation from research to prototype, to development and commercialization, and finally to deployment and evaluation (and back to research). The differing, non-overlapping missions and reward systems built into each agency make it too easy to lose momentum at any of these stages. Longer funding cycles are needed to elicit requirements and integrate these into a research agenda, and then enter into an "agile" software cycle of develop, evaluate, and deploy. One year is barely adequate for the fist step (eliciting requirements, understanding the domain, and setting up a collaboration); three years is more adequate to developing and evaluating tools with decision maker collaborators; special two year supplements for deployment (given prior evaluation) would continue a cycle of innovation.

   Other questions remain in terms of sustaining this research: How would researchers funded by NSF find collaborators in the field so they can best understand in adequate detail resource problems, extract the research issues, and test their prototypes? How might research results and prototypes funded by NSF make their way to resource managers in the form of information technology deployed in field offices? How would the evaluation of new products, and an understanding of their strengths and weaknesses, be fed back into this loop to inform new research?

[I'd like us to think on #4 a bit, and come forward with some specific recommendations. – jbc]