# 2003 National Conference on Digital Government Research
## Panel : Biodiversity and Ecosystem Informatics (BDEI)

Judy Cushing (Moderator), The Evergreen State College (judyc@evergreen.edu)
Kate Beard-Tisdale, U. of Maine; Barbara Eckman (IBM); Eric Landis, Natural Resources Information Management; John Schnase, NASA; Rob Stevenson, U. of Mass. (Boston)

In June, 2000, an NSF-NASA-USGS sponsored workshop brought biologists, ecologists and resource managers together with computer scientists to identify the CS/IT research issues that impede biodiversity and ecosystem research and ecosystem information infrastructure. Their report sparked the National Science Foundation to issue a call for proposals[1] and 15 awards were made that fall. In February, 2003, Principal Investigators of Biodiversity and Ecosystem Informatics (BDEI-CISE-EIA) planning and incubation grants came together with agency representatives (NSF, USGS and NASA) and other interested parties to report research results and refine the BDEI research agenda[2].

**Workshop participants emphasized that wide ranging domain expertise and connections to climate, agriculture, resource management, recreation, and public health make ecosystem informatics a high priority for society, and strongly encourage research in the following research areas:**

- **Terminology management with ontologies and glossaries,**
- **Data provenance, metadata, and annotation,**
- **Adaptive, flexible database schemas and schema management,**
- **Support for spatio-temporal data and other domain-specific data types,**
- **Uncertainty management and data cleaning,**
- **Digitization and processing of museum specimens, and**
- **Management of mathematical and statistical models (with data) and modeling support.**

BDEI research projects (and ensuing research areas) were categorized into four topic areas: 1) Semantic Data Integration, 2) Spatio-Temporal Data / Remote Sensing, 3) Modeling and Forecasting, and 4) Putting it into Practice: Domain Research Practices, Community, Workflow. The DGO BDEI panel organized the major research results of the BDEI planning and incubation grants into these four categories. One panelist also addressed overlap of the BDEI research agenda with that developed at an National Library of Medicine (NLM) sponsored Workshop on Data Management for Molecular and Cell Biology[3]. A brief summary of each panelist's remarks follows, organized by the four topic areas. Where the panelist presenting the subgroup's results was not the chair of the BDEI subgroup, his or her name is also given.

**Semantic Data Integration (K. Beard).** Data needed to address critical questions in ecology are scattered, heterogeneous, and complex. Large volumes of diverse data types must be semantically integrated to support research. Early integration technologies generally assumed that terms and formats (syntax) and meaning of terms (semantics) used across data sources were the same. In the web environment and the decentralized, heterogeneous collections of information relevant to ecological research assumptions similar semantics do not hold. Research approaches include thesauri and ontologies, but metadata services (in particular for taxonomies) are also critical. Existing digital gazetteers (USGS GNIS and DMA) are a good start, but typically lack spatial definition and ecology-specific data structures; agent architectures show promise for linking ontologies with various metadata services.

**Spatio-temporal Data & Sensing Technologies (J. Cushing and G. Henebry).** We now face multiple biological and ecological threats that are occurring at diverse tempos, e.g., West Nile virus, invasive species,

---

[1] See NSF Dear Colleagues Letter, available at http://bio.gsfc.nasa.gov/bddcl.pdf
[2] Reports and presentations of both the 2000 and 2003 BDEI workshops can be found at http://canopy.evergreen.edu/bdeipi
[3] See **http://www.lbl.gov/~olken/wdmbio**.

drought, biological warfare agents, global change. At the same time, new sensors – aloft, among, and in situ – yield more and more data. A recent confluence of sensing technologies – orbital sensors complemented by mobile teams of citizen-scientists and specialists equipped with geo-referential proximal sensors and robust wireless networks of reactive sensing agents – result in data streams rich in dimensionality and across a wide range of spatial, temporal, spectral, radiometric, thematic, and taxonomic scales. Robust representations that capture the complexities of environmental patterns and processes, are required to exploit these data streams. Flexible Database schemata and sophisticated queries, algorithms for spatio-temporal analysis, wireless reactive agent networks and tools for metadata acquisition, management and interpretation were cited as the most critical research areas. In addition enhanced sensory presentation of environmental data encompassing visual, sonic, and haptic feedback systems were seem as needed to explore high-dimensional data spaces.

**Modeling and Forecasting (J. Schnase and J. Clark).** The modeling and forecasting group identified five research challenges. 1) Advanced frameworks, including hardware for adaptive and intelligent systems and high performance computing needs. Memory and speed continue as important bottlenecks, so coupling of models, modular models, and hardware grid /distributed computing are needed. 2) Model-Data Interaction. Many ecological problems could be informed by existing large data sets, but methods are needed for inference of high dimensional problems. Data are indirect, massively unbalanced with missing observations, and subject to many sources of stochasticity. Processes interact at a range of scales and frequently hidden, 'parameters' are often more like variables, and data exhibit variability, uncertainty, and complexity. New approaches are needed for sensor networks. 3) Visualization Tools. More than just "pictures", meaningful visualization must be part of working models on which measurements may be made and change simulated. 4) Software Developments. Biodiversity science requires more efficient algorithms, ways to deal with incomplete knowledge, better understanding of spanning spatiotemporal scales, and techniques for high dimensional problems. 5) Information Products. Derived data products and provisioning data at multiple places were viewed as key.

**Putting it into Practice: Domain Research Practices, Community, Workflow (R. Stevenson).** Because the promise of advancement hinges on closing various "digital divides", shaping an ecoinformatics culture and developing socio-technological partnerships, this team focused on the various "digital divides" and on social and ethical conundrums among various players: industry and science; bioinformatics and ecosystem informatics; ecologists conducing "big environmental science" and traditional individual ecologists working in relative isolation; scientists and information managers; professional and citizen scientists; and the tripartite scientists, resource managers and policy makers. Because R&D budgets of telecommunications and consumer electronics companies as well as the defense and the health services industries, and even bioinformatics, dwarf those of institutions working on environmental informatics, BDEI should look there for technologies to adapt and transfer, keeping in mind certain special if not unique requirements. Even once these problems are solved, however, extensive training will be required. Stakeholders should adapt open source models to environmental sciences enabling technologies, in particular for specialized robots to digitize museum collections. This group emphasized that just creating the technology is not enough, and that social and ethical components abound: who, for example, should contribute information, who should have access and how information should be shared, and how many research resources should be allocated for information management?

In addition to the above reports on the BDEI research agenda, DGO panelists Eric Landis and Barbara Eckman presented (respectively) the importance of data archiving within the Forest Service and National Parks, and the overlap between research agendas of BDEI and bioinformatics. The primary differences between the two areas were seen to be (in ecosystem informatics) a stronger emphasis on sensor networks and modeling, along with a dire need to process (digitize) museum specimen data. In molecular biology, problems arising from propagating uncertainty in derived data products seemed to have significantly stronger emphasis. Also, the two areas might rate major research topics differently.

For more information about the Biodiversity and Ecosystem Informatics Workshops, contact Judy Cushing (judyc@evergreen.edu) or go to http://canopy.evergreen.edu/bdeipi