

Eco-Informatics for Decision Makers Advancing a Research Agenda

Judith Bayard Cushing

The Evergreen State College, Olympia, WA, USA

judyc@evergreen.edu

Tyrone Wilson

National Biological Information Infrastructure, U.S. Geological Survey, Reston, VA, USA

tyrone_wilson@usgs.gov

Larry Brandt, Valerie Gregg, Sylvia Spengler (National Science Foundation)

Alan Borning (University of Washington), Lois Delcambre (Portland State University),

Geoff Bowker (Santa Clara University), Mike Frame (USGS/NBII),

János Fülöp (Hungarian Academy of Sciences),

Carol Hert (University of Washington-Tacoma),

Eduard Hovy (University of Southern California), Julia Jones (Oregon State University),

Eric Landis (Natural Resources Information Management), John L. Schnase (NASA),

Charles Schweik (University of Massachusetts-Amherst), William Sonntag (EPA)

Abstract. Resource managers often face significant information technology (IT) problems when integrating ecological or environmental information to make decisions. At a workshop sponsored by the NSF and USGS in December 2004, university researchers, natural resource managers, and information managers met to articulate IT problems facing ecology and environmental decision makers. Decision making IT problems were identified in five areas: 1) policy, 2) data presentation, 3) data gaps, 4) tools, and 5) indicators. To alleviate those problems, workshop participants recommended specific informatics research in modeling and simulation, data quality, information integration and ontologies, and social and human aspects. This paper reports the workshop findings, and briefly compares these with research that traditionally falls under the emerging eco-informatics rubric.

1 Introduction

Decision makers at all levels of government and at NGOs who manage natural resources or carry out ecological or environmental policy face significant information technology (IT) problems when integrating ecological or environmental information. Ecology and environmental decision makers work with information providers and data managers, and seek a wide variety of information sources, but little of the data used to

produce these sources is collected specifically for the decision making at hand. Thus, the decision maker is faced with data gaps, data presentation mismatches, and finding appropriate indicators. These IT issues suggest computer science research needs in information integration, modeling and simulation, data quality, and human-centered issues such as training, technology transfer, best practices for information provision and use, and human-friendly software. While a growing body of research has focused on information technology to help ecology researchers [pcast,bdei], solving IT problems in natural resource management is not simply a matter of adopting the technology developed for other domains, or even applying research completed under the eco-informatics rubric to the decision-makers' problem space. As suggested in discussions of eco-informatics at a Digital Government Conference, the problems are different, and the field of eco-informatics should be extended to include helping decision makers (e.g., policy makers and natural resource managers) utilize data and information more effectively. [dgo.04]

Eco-informatics problems faced by natural resource decision makers require, in addition to new research, sustaining innovation in the public and NGO sectors just as in digital government research [dgo.05]. Researchers must find the right domain problem, distill research that will prove fruitful to a range of stakeholders, find the right agency collaborators, and manage everybody's expectations. [hovy] Researchers in eco-informatics decision-making must also consider combining quantitative with qualitative information and have a basic understanding of decision making. If computer scientists and social scientists in the academy are not prepared to take on these challenges in addition to demanding research, natural resource eco-informatics will lag behind informatics in other science and policy domains. We base this assessment on the complexity introduced by public policy requirements added to already complex scientific eco-informatics issues.

2 Eco-informatics Problem Space for Natural Resource Management

Eco-informatics is about both biodiversity-rich conservation managed systems and natural resource protection and human health impacts of environmental, anthropogenic pollutants, on the other. Rather than sorting out different informatics needs for these two areas, we recognized that the latter presupposes a command of the former, and focused on non-human-health-centered ecological constituencies. Another perspective can be found in Europe, where the research is much broader in nature and includes health and security, as well as ecosystem function. [jensen]

To map the problem space for natural resource management eco-informatics, we note that many organizations have developed IT for natural resource decision makers, and we laid out the eco-informatics problem space collaboratively with personnel from the USGS, NASA, EPA, State Agencies and InterState Consortia. [e.g., guldin, sugarbaker] Resource management informatics is hard, and data and tools form a demand cycle; the more successful one is, the more demand arises. Exemplary projects, as at Ohio State and Oregon State universities show how researchers and coastal policy makers might collaborate, and exemplify the research complexity. Coasts are interaction zones of land, sea, and air, and although they occupy only 3% of sea surface area and 0.5% of ocean

volume, about 70% of global fish resources spend time there. Further, about 60% of the world's human population resides close to the coast zone, exploiting it for food, recreation, transport, waste disposal, etc. This brings materials discharge from land to coast, and causes environmental changes through physical, chemical, and biological processes. Enhanced data handling capacity and cooperation among intergovernmental agencies are essential for integrating required multi-source data. [li, wright] Drawing on these and other experiences, we categorized the problem space for natural resource management eco-informatics into: policy, data presentation, data gaps, tools, and indicators. [context]

2.1 Policy Problems

Policy issues related to ecological and environmental information and decision-making include, but are not limited to, problems that organizations encounter because of policies related to: 1) the provision, production and maintenance of eco-informatics tools and information; 2) the use and possible abuse of tools and information; 3) the cross-organizational sharing (or lack thereof) of tools and information; and 4) the communication (or lack thereof) of environmental management decisions grounded upon eco-informatics-based analysis. From here on, “BDEI” (biodiversity and eco-system informatics) will be used interchangeably with eco-informatics.

1. Provision, production, and maintenance, e.g., data curation or archiving. BDEI tools and information must begin with user needs assessment, but in some cases developers are not doing this. We recommend research into why this is so, how the problem differs from other IT application areas, and how to solve the problem. Though costly and inefficient, data are sometimes collected but only used once, or even not at all. Better systems of metadata and storage retrieval might ensure that what data are collected or generated are used and shared. Whether BDEI tools and information should be treated as public, private or toll goods is another issue. While the tools and information could be considered public goods, there might be important reasons to treat them as a private or toll good in order to collect revenue to absorb some costs.

2. Use and possible abuse of tools or information. Issues here include translating from uncertain scientific models to policy decisions that require a legal burden of proof, and determining whether scientific evidence justifies a policy decision when there is uncertainty in the data. Tools that might be useful for policy analysis are not used in decision making because they a) take too long compared to the political cycle; b) cost too much; c) are based on unrealistic assumptions; or d) are too complex or technical.

3. Cross-organizational data and tool sharing, or lack thereof. We note two levels of organizational cooperation: sharing tools or information, and co-production of tools or information. Organizations generally avoid paying to develop tools if another organization is also involved – a classic free rider problem in collective action theory. Further, organization policies can be barriers to co-production, if employees are not recognized for such efforts. Where organizations are interested in sharing information or models, they may be hindered by inadequate metadata or ontologies that would allow integration with their own data and models. We see both carrot and stick approaches for encouraging data

owners to produce and maintain metadata, e.g., employee performance rewards or positive recognition, or mandates with negative consequences for noncompliance.

4. Communication (or lack thereof) of environmental management decisions grounded in informatics-based analysis. Organizations sometimes do a poor job communicating issues discovered through BDEI analysis – an information diffusion problem involving mass communication. Organizations or policy-makers are sometimes caught off-guard by an environmental or ecological problem, and face difficulties addressing the problem because little or no data exist. Further the extant data might exhibit a linear trend when in fact the pattern is more complex. These problems are illustrative of decision-making or policy under circumstances involving uncertainty. [policy]

2.2 Data Presentation

Data Presentation problems arise from complex interactions between user needs and data (e.g., metadata, raw data, information, accuracy specifications, methods, documentation, policy). System limitations (e.g., software modalities, hardware availability or costs); and information format further complicate presentation. Needed 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. This problem area can be distilled into two major components: 1) presentation as the mediator between users and their needs, and between task and data or metadata and their characteristics, and 2) the set of research questions and themes that relate to the facilitation of that mediation role. Presentation options must reflect the user experience as well as the nature of the data, with constraints. Presentation types may need to reflect a number of user dimensions: 1) User needs, perhaps conceptualized as tasks or time available, or context, and 2) User characteristics, including preferences, (dis)abilities, and computing capabilities available. On the data side, presentation could reflect the nature and amount of data, metadata available, data and metadata quality, data preparation activities, and policies such as privacy and confidentiality. Presentation instantiations and approaches need to reflect the marriage of the user and the data. In addition, presentation media add their own affordances and issues. Different software modalities may have different suitabilities for different data types, and different hardware media have different costs, availability, and permanence. These components will suggest a range of research questions that will help understand presentation for BDEI decision making. [presentation]

2.3 Data Gaps

Geographic data gaps between biodiversity-rich and conservation-managed land areas adversely impact decision making. These problems stem from lack of the needed data sets or access to them, disjoint data sets that require manipulation to compensate for temporal or spatial gaps, an emphasis on adaptive management which outpaces data reliability, or a paucity of database professionals upon which resource managers can call. Major issues include how to appropriately generalize fine-scale data that will necessarily contain gaps, and decision makers' and policy makers' sensitivity to uncertainty. Next

steps to refine this problem area would be to address the original data needs, and define review criteria, such as stable standards for data collection and documentation.

2.4 Tools

Major BDEI tool problems apply across the board to scientific informatics research, namely, how one balances longer term research to advance functionality with supporting users in the short term. Tool problems involve 1) a lack of a tool “clearing-house”, i.e., from the developer side getting a tool out to users, and from the user side finding and evaluating tools and determining if a given tool can be applied to other problems or with other data than what it was developed for. 2) new or different data types and data collection methods, 3) lack of user frameworks, product suites, and development standards, 3) lack of tools to support metadata issues, and 4) social science issues of usage, sharing, and adoption.

2.5 Indicators

Indicator problems exist because indicator definition, relevance, and value are neither well-defined nor communicated. Constituents may be uneasy with environmental measures, and data gaps effect reliability and trust that these stakeholders have in indicators. Finally, the inherent complexity of the ecosystem further complicates this issue. Prime examples of the complexity that arises in using indicators include the Death Valley Pupfish and the Washington State Shellfish Bed Closures.

3 Research Issues

Teasing out research issues from the natural resource management problem space was a four-step process. 1) We examined three current research projects to see how interdisciplinary approaches and government partners were involved. 2) In breakout sessions, we articulated research issues, which were 3) in plenary sessions critiqued by a panel of resource managers and researchers. 4) Finally, breakout groups reconvened to refine and prioritize issues, identify strategies for sustaining research, and suggest resource management case histories that exemplified the need for research. [research]

Three NSF Digital Government research case studies, the Forest Portal, UrbanSim, and Understanding Government Statistics. [blm, borning, hert], gave an initial set of research issues to consider. The Forest Portal, an adaptive management tool that harvests information to sustain forests, highlighted the importance of collaboration between federal agencies and academic institutions, and demonstrated the capabilities of using metadata attachments. UrbanSim demonstrated how ecological models and partnerships contribute to data collection, preparation, and assessment, which in turn would likely lead to realistic policy scenarios and major policy applications. The GovStat project models user access to U.S. government statistical information to better integrate data across agencies. The project emphasized deploying prototypes to identify research challenges and designing an

interface that relies on metadata generated from the web sites. Research issues were categorized into four major areas: 1) modeling and simulation, 2) data quality, 3) data integration and ontologies, and 4) social and human aspects.

3.1 Modeling and Simulation

Modeling and simulation research issues included: coupling diverse models, addressing values in design for diverse stakeholders, incorporating new visualizations for results, representing error and uncertainty when presenting information to decision makers, challenges in handling large data sets, and open source modeling infrastructure. This group emphasized proposed that an open-source, flexible, reusable modeling infrastructure, along with the social practices that sustain it would allow researchers and decision makers to experiment freely with new models or change existing ones.

3.2 Data Quality

Data quality research issues primarily involve how to determine and communicate uncertainty to decision-makers who 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. The overarching research question invoked by this problem is the extent to which uncertainty associated with data quality and synthesis really has an influence on policymaking and plan implementation. Research issues arise also in individual studies and data sharing where diverse data sources are combined. Research is needed to develop methods for 1) reducing the introduction of error when datasets are created and combined, measuring and logging error at each stage of the study and 3) characterizing relationships among errors – additive, multiplicative, averaging.

Where data are shared, for example in data harvesters such as the Long-term Ecological Research network's Clim-DB and Hydro-DB, the major issue is the extent to which metadata can become an integral part of the dataset. What happens to metadata when multiple sources are integrated? How can metadata management be automated? How can data standardization help in combining metadata from multiple sources? How could metadata become a part of the data set. The research challenge is how general can the tools that manage data quality become, and whether they be applied to a wide range of ecological datasets. NSF could develop and publish metadata standards across all grants, instead of just for certain programs. Successful metadata efforts include the Federal Geographic Data Committee within the USGS and the LTER information manager standards used in internal reviews of LTER projects.

To determine the influence of uncertainty associated with data synthesis on policymaking and plan implementation, studies could be done of decision-makers perceptions of the value of science findings made from synthesized or integrated data. For example, data harvesters such as Clim-DB and Hydro-DB have generated publications from combined datasets, which are perhaps being used by land managers or decision makers in the Forest Service and NOAA. How is the increase in power associated with

data synthesis balanced by the increase in uncertainty associated with the ways in which the errors were combined? How might synthetic studies stand up in courts of law in comparison with other forms of “expert testimony?”

3.3 Information Integration and Ontologies

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, tools to support data integration, and evaluating 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 ontologies in ways that go beyond simple versioning. The semantics of BDEI is critical, and includes defining and operationalizing meanings, data standardization, and semantic services. Transferring knowledge from other domains to BDEI is itself research. Quality control, data access, and collaborative decision making support are also critical.

The reclassification of rainbow trout as salmon in the early 1990s and a subsequently implemented information system had broad-reaching effects; the moral being that no indicator is innocent, and IT systems have social consequences. How can computer scientists and developers be sensitized to the value judgments inherent in data collection, ontology generation, and modeling? Future IT applications should warn scientists and policy makers of impending circumstances. [ontology]

3.4 Social and Human Aspects

Research issues identified in this area included: 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, and user requirements and system design. 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 to identify habitat parcels most threatened by human encroachment. Ideally, Jane would like policy-relevant modeling to 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. We considered this scenario 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 to integrate information over time and across agencies, evaluate legacy data, identify indicators, visualize alternative actions, and model current ecological conditions. [social]

4 Conclusions, Broader Impacts, and Recommendations.

One metaphor useful in understanding our natural resource management vision is a fictitious, ideal decision making tool, dubbed Yoda. Yoda sees decision makers as those who choose among alternatives, and what they do as integrating information – via sharable data structures, compatible software tools, human collaboration, and understanding outcomes. Theirs 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 in this report dictate that it will take decades to solve the problems – all while species and ecosystems disappear at an increasing rate, and natural resources are depleted. Thus, we need to prioritize the critical informatics problems – ask where problems intersect across agencies and environments to find the greatest synergies, which of those with the greatest intellectual merit could be solved with focused R&D, and where public and private funds could be leveraged. A follow-on workshop of resource managers, eco-informatics professionals and computer scientists, itself followed by online surveys, auctions or futures markets could accomplish this. Because problems are both technical and sociological, a few well-chosen broad projects in those areas could serve other more focused research.

Two unanticipated issues emerged as we followed our agenda: 1) feedback loops and 2) the nature of decision making. If resource managers become more effective, the effect on managed system is not negligible. We also saw that understanding decision making is critical for doing the work. Environmental issues are complex with considerable uncertainties, but in the political and policy arena many decisions are placed in a “yes or no” context. Thus, one workshop product is a decision making primer. [decision]

Communication enables collaboration (human centeredness), trustworthiness (ontologies), and data sharing (data integration). Social science is characterized by indigenous local and community knowledge plus the 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. One real challenge of this area is the difficulty involved in pursuing research in one of these areas without at least some understanding of the others.

Another challenge involves training computer scientists and social scientists to work in eco-informatics and natural resource management. A graduate student team considered how researchers might articulate educational impacts, involve students in research, and use research as a teaching tool. The students saw the ethical issues around large data repositories as a particularly fruitful area for teachable moments, and funding interdisciplinary mission-oriented tasks that force addressing local problems as a way to

pursue these goals. The students encourage NSF to partner with agencies that support applied student research. [gradStudent] We further encourage the early focus in eco-informatics and decision making to be on ecological and biodiversity issues, as environmental health decision making is even more complex and requires natural resource management as input.

Funding agencies must work together and with principal investigators, information managers and decision makers to sustain and encourage innovation in this area. How would researchers funded by NSF find collaborators so they can best understand resource problems, extract the research issues, and test prototypes? How might research results and prototypes make their way to resource managers deployed in field offices? How might product evaluation be fed back to inform new research? 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 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 first 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, and we recommend special two-year supplements for deployment (given prior evaluation) would continue a cycle of innovation.

Finally, considerable attention must be paid to constant re-prioritization of the research agenda, and assuring development of tools that promise, through extensibility, applicability to a wide range of problems, as they arise in important eco-systems. We emphasize the importance of keeping a range of research projects in the pipeline – from highly theoretical and generalizable, to working prototypes developed by researchers and resource managers, to deployment experiments.

5 URLs Cited in this Report

dg.o 2005 Panel, Digital Government and the Academy (Delcambre, Giuliano), May 16-18, 2005. <http://dgrc.org/dgo2005>.

dg.o 2004 Birds of a Feather (Schweis et al), May 24-26, 2004.
http://dgrc.org/dgo2004/disc/bofs/bof_ecoinformatics.pdf

PCAST Panel on Biodiversity and Ecosystems, “Teaming with Life: Investing in Science to Understand and Use America’s Living Capital”, March 1998.
<http://clinton3.nara.gov/WH/EOP/OSTP/Environment/html/teamingcover.html>.

BDEI - Biodiversity and Eco-System Informatics Workshops

BDEI-1. NSF, USGS, NASA Workshop (Maier, Landis, Cushing, Frondorf, Silberschatz, Frame, Schnase), NASA (Goddard), June 2000. <https://www.evergreen.edu/bdei/2001/>

BDEI-2. PI’s Meeting Report (Cushing, Beard-Tisdale, Bergen, Clark, Eckman, Henebry, Landis, Maier, Schnase, Stevenson), NSF (Arlington), February 10, 2003.
<https://www.evergreen.edu/bdei/2003/>

BDEI-3. Eco-Informatics for Decision Making (Cushing, Wilson, et al), The Evergreen State College, December 13-15, 2004. <http://www.evergreen.edu/bdei/blm> <http://www.cse.ogi.edu/forest/papers/blm-briefing.ppt>
borning <http://www.evergreen.edu/bdei/presentations/borning.pdf>
context http://www.evergreen.edu/bdei/presentations/tuesbreakout1_combined.pdf
decision <http://www.evergreen.edu/bdei/decisionMaking>
gradStudent <http://www.evergreen.edu/bdei/presentations/gradStudentDraft.pdf>
hert <http://www.evergreen.edu/bdei/presentations/hert-tuesdaylunch1.pdf>
hovy <http://www.evergreen.edu/bdei/presentations/hovy.pdf>
jenson <http://www.evergreen.edu/bdei/presentations/jenson.pdf>
li <http://www.evergreen.edu/bdei/presentations/li.pdf>
ontology <http://www.evergreen.edu/bdei/presentations/wedbo3ontology.pdf>
presentation <http://www.evergreen.edu/bdei/presentations/wedbo3summary.pdf>
policy <http://www.evergreen.edu/bdei/presentations/summaryPolicygroupfinal.pdf>
research http://www.evergreen.edu/bdei/presentations/tuesbreakout2_combined.pdf
and http://www.evergreen.edu/bdei/presentations/breakout3_combined.pdf
social <http://www.evergreen.edu/bdei/presentations/wedbo3humancenterednessb.doc>
sugarbaker <http://www.evergreen.edu/bdei/presentations/sugarbaker.pdf>
tosta <http://www.evergreen.edu/bdei/presentations/GreybeardNT.pdf>
wright <http://www.evergreen.edu/bdei/presentations/wright.pdf>

Acknowledgements

The authors thank those who presented research, development and needs assessment at the workshop, all of which informed this report: Frank Biasi (Nature Conservancy), Alan Borning (Univ. Washington), Geoffrey Bowker (Santa Clara Univ.), C. Marie Denn (USPS), Richard Guldin (US Forest Service), Carol Hert (Univ. Washington-Tacoma), Eduard Hovy (Univ. S. California), Stephen Jensen (European Union), Paul Klarin (State of Oregon), Ron Li (Ohio State Univ.), Molly O'Neil (State EPA Network), Phil Rossignol (Oregon State Univ.) John L. Schnase (NASA), Mark Simonson (Puget Sound Regional Council), Larry Sugarbaker (NatureServe), William Sonntag, Nancy Tosta (Ross & Assoc Environmental Consulting), Dawn Wright (Oregon State Univ.), and to János Fülöp, David Roth and Charles Schweik (Univ. Mass-Amherst) and G. Bowker for writing the Decision Making in Eco-Informatics Reports.

We also thank other workshop participants: Bill Backous (Wa. State Dept. of Ecology), Bruce Bargmeyer (University of California, Berkeley), Thomas Beard (USGS), Clifford Duke (Ecological Society of America), Kevin Gergely (USGS), Vivian Hutchison (USGS), Brand Nieman (EPA), Craig Palmer (Univ. Nevada-Las Vegas), Pasky Pascual (USGS), Sherry Pittam (Oregon State Univ.), Louis Sweeny (EPA), Tim Tolle (USFS-Ret.), Andrew Wilson (USDA), Steve Young (EPA), as well as the graduate student panel Steve Abercrombie, Gigi Sanchez, and Hiroshi Sato, and Canopy Database Project staff Anne Fiala, Aaron Crosland, Chris Hardy, and Mike Finch (all of The Evergreen State College) who provided excellent local support. Aaron Ellison of Harvard University and Elaine Hoagland of The National Council for Science and the Environment for their very helpful comments on an initial short version of this report.

We acknowledge financial and administrative support from the National Science Foundation (NSF IIS 0505790), and emphasize that the opinions stated in this paper are not necessarily those of NSF.