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),
John L. Schnase (NASA), William Sonntag (EPA),
Janos Fulop (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),
Charles Schweik (University of Massachusetts-Amherst)

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

The informatics tools needed to solve environmental challenges (e.g., global climate change, emerging diseases, decreasing biodiversity, and waning resources) are currently being researched and developed under the rubric of eco-informatics. These needs were characterized in the 1998 PCAST report that 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, which is available at http://clinton3.nara.gov/WH/EOP/OSTP/Environment/html/teamingcover.html.

Most eco-informatics research efforts subsequent to PCAST, and as articulated by researchers and agency representatives at workshops[1] sponsored by the National Science Foundation (NSF), National Aeronautics and Space Administration (NASA) and National Biological Information Infrastructure (NBII), have focused on tools to help increase research productivity and increase the published availability of research data. A growing body of research has emanated from this and similar efforts.

In mid 2004, researchers at an NSF Digital Government Conference[2] participated in a lively conversation on future directions of eco-informatics. They extended the eco-informatics vision to include needs of decision makers (e.g., policy makers and natural resource managers) in utilizing eco-informatics products more effectively. They recommended that NSF and USGS fund a workshop devoted to decision-maker needs regarding eco-informatics. Thus, the December 2004 workshop focused specifically on informatics tools to support ecological and environmental decision makers.

Information technology is critical to natural resource managers. For example, EPA science advisors in gauging pesticide exposure risk consider that CS/IT software could potentially avoid protracted disputes regarding assumptions about public health risks (reported in *The Washington Post*, October 21, 2004). At this point however the potential is far from being attained. Potential decision makers at all levels of government and at NGOs who manage natural resources or carry out ecological or environmental policy often face significant information technology (IT) problems when integrating ecological or environmental information, and non-governmental organizations have similar goals and problems. 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 (often indirectly) with many information technology (IT) issues, including data gaps, data presentation, and how to use or create 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.

Eco-informatics problems faced by natural resource decision makers require, in addition to new research, particular efforts to sustain and encourage innovation, research and development in the public and NGO sectors. These findings are not unlike needs articulated by digital government researchers (for example, at a May 2005 Panel at dg.o

---

[1] See www.evergreen.edu/bdei.

[2] National Conference on Digital Government Research, http://www.dgrc.org/dgo2004, Eco-Informatics BoF Participants: Chaitan Baru, Judith Cushing, Stefan Falke, Mike Frame, Bill Hodgkiss, Eric Landis, Maria Matevosyan, Peter McCartney, G.P. Patil, Jon Schweiss, Sharon Shin, William Sonntag, Sylvia Spengler, Charles Taillie, Bill Waltman, Jessie Wilbur, Tyrone Wilson. Special thanks to Val Gregg (NSF), Sue Stendebach (EPA) and Bruce Bargmeyer (LLBL), who contributed to formulating the agenda.

2005 organized by Lois Delcambre and Gen Guiliano, http://dgrc.org/dgo2005). Eduard Hovy, calling upon his experience with eco-informatics projects funded by the NSF digital government program, notes that considerable attention must be paid to finding the right domain problem on which to focus, distilling a range of research that will prove fruitful to a range of stakeholders, finding the right agency collaborators and then managing expectations (https// www.evergreen.edu/bdei/presentations/hovy.pdf).

Researchers in this area must consider combining quantitative with qualitative information, and have a basic understanding of decision making. Like other scientific computing research, the field would benefit greatly from considerable open-source, flexible infrastructure (such as a reusable modeling infrastructure), along with the social practices that would sustain it. If computer scientists and social scientists in the academy are not prepared to take on these challenges in addition to demanding research, workshop participants believe that natural resource eco-informatics would continue to lag considerably behind informatics in other science and policy domains. This prediction is based on the complexity introduced by public policy requirements added to already complex scientific informatics issues; solving these problems is not simply a matter of adopting technology developed for another domain.

2 Ecoinformatics Problem Space for Natural Resource Management

Eco-informatics is about both biodiversity-rich conservation managed systems and natural resource protection, on the one hand, and about human health impacts of environmental, anthropogenic pollutants, on the other. Rather than sorting out different informatics needs for these two separate areas, the workshop recognized that the latter area likely presupposes a command of the former, and thus focused primarily 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. (https:// www.evergreen.edu/bdei/presentations/jensen.pdf.)

To map the problem space for natural resource management eco-informatics, we note that many organizations now have considerable experience developing information technology for natural resource decision makers. At the workshop, the context of what we call the eco-informatics problem space was laid out by representatives from the U.S. Geological Survey, Environmental Protection Agency, U. S. Forest Service, NASA, State Agencies and InterState Consortia such as State/EPA Environmental Information Exchange Network (ECOS). Policy makers and their clients' information needs form communities of interest as well as of place, and thus the required information technology is very broad. An example of a community of place is the ecology of Mt. Rainier, which is a well-loved, locally used, park in the Pacific Northwest. One can easily see, however, both communities of place and interest with regard to the Escalante Natural Monument, which might be valued locally for grazing but nationally for its special scenery. The possible political aspects of certain data make equitable information access and mass customization necessary, as well as clearly distinguishing among measurements, indicators, and interpretations. Finally, metadata and validation are all essential to biodiversity and ecosystem (BDEI) decision making. As Rich Guldin stated in his presentation: "Better data lead to better dialogue, which leads to better decisions."

As Larry Sugarbaker of NatureServe observed, projects at non-governmental agencies have found that conservation informatics is hard, and data and tools form a demand cycle; the more successful one is, the more demand arises. Biodiversity data management and collection would be more efficient were data managed in common formats, with better decision support tools, such as a common framework for geographic information. Requirements for common frameworks, however, as we know from other scientific application areas, come with their own sets of problems.

In spite of these contextual challenges, however, some successful exemplary projects demonstrate how current and future eco-informatics might be used to engender better resource management. A recently deployed collaborative project between the EPA and The Environmental Council of the States (ECOS) 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 has been a key first application of this technology. As evidence of the perceived value of this collaboration, many states have surprisingly been ready with data even earlier than the jointly agreed upon deadline. NASA's Science Mission experience with decision support for earth science, in particular for the invasive species project, shows how remotely sensed raw data (observations) can now be used in conjunction with models (predictions) as input to decision support tools.

Two NSF projects, at Ohio State University and Oregon State University, demonstrate successful collaboration between university researchers and coastal policy makers, and exemplify the complexity of informatics problems. The coastal zone is an interaction zone of land, sea, and air. Although the coastal zone occupies only 3% of all the sea surface area and 0.5% of ocean volume, about 70% of global fish resources spend part of their life in the coastal area. About 60% of the world's human population resides close to the coast zone, which is exploited by humans for food, recreation, transport, waste disposal, and other needs. With the increase of the human activities, many materials discharged from the land, spread to the coastal area and cause environmental changes through various physical, chemical, and biological processes.

Excessive discharge and uncontrolled human activities in the coastal zone creates environmental problems, such as habitat modification and destruction, and ocean pollution. For instance, excessive loading of nutrients from watershed and various dredging operations and shoreline developments during the 1950s, '60s and '70s resulted in about 70% loss of seagrass beds in Tampa Bay, FL. Research indicates that the factors controlling seagrass distribution include nutrient loading, water quality, light, water depth, tide and water movement, salinity, temperature, climate change, and anthropogenic impacts. To facilitate sustainable coastal management, an important focus is to investigate the natural variability of coastal ecosystems and the complex interactions among biological and physical systems in coastal environments. . Research integrating remote sensing techniques and three-dimensional conceptual and quantitative models is needed to explore physical, chemical, and biological processes in coastal environments.

Both enhanced data handling capacity and cooperation among intergovernmental agencies are essential to integrating these multi-source bio-chemical and geo-spatial information. In Oregon and Tampa Bay, community partnerships have yielded some success with hazards management, watershed assessment, and ocean protection, where decisions depend on accurate resource status information, but have been even more successful at highlighting the research needs needed to monitor coastal eco-environmental changes and predict future impacts and possible hazards. More information about these case studies can be found on the workshop website.

Drawing on these and other partial successes, workshop participants categorized the problem space for natural resource management eco-informatics as problems in the following areas: policy, data presentation, data gaps, tools, and indicators. See https://www.evergreen.edu/bdei/presentations/ tuesbreakout1_combined.pdf for the workshop presentations of the five discussion groups.

2.1 Policy Problems

Areas for consideration as part of the policy problem space related to ecological and environmental information and decision-making include (but are not limited to) problems that organizations (across all sectors, public, private and nonprofit) encounter because of their policies related to: 1) the provision, production and maintenance of eco-informatics tools and information; 2) the use (and possible abuse) of eco-informatics tools and information; 3) the cross-organizational sharing (or lack thereof) of eco-informatics tools and information; and 4) the communication (or lack thereof) of environmental management decisions grounded upon eco-informatics-based analysis.

1. Provision (e.g., financing), production, and maintenance (e.g. data curation or archiving). The production of EI tools and information need to begin with an understanding of user needs, but in some cases, developers of tools or information are not doing this. Short term research into why this is the case, and how the problem differs from all other IT application areas, and how to solve the problem, would be beneficial. It is relatively common-place, though costly and inefficient, that lots of data is collected but only used once (or in some instances, not at all). There is a need for better systems of metadata/storage retrieval systems to ensure that what data are collected or generated are used (and shared) more frequently. Organizations face a policy dilemma related to what kind of good to treat EI tools and information. Should they be treated as public, private or toll goods? On the one hand, EI tools and information could be considered public goods for they have the potential to be used by others outside of the organization. At the same time, there might be important reasons to treat them as a private or toll good, in order to collect some revenue to absorb some of the costs to produce this information.

2. Use (and possible abuse) of eco-informatics tools or information. For example, how do you translate from uncertain scientific models to policy decisions that require a legal burden of proof? In other words, how do you determine whether scientific evidence is enough and defensible to justify a policy decision when there is uncertainty in that data? One participant noted that advances in this area include, for example, Bayesian statistical

approaches. In some circumstances, EI tools that might be useful for policy analysis may not be used by decision-makers (or more probably their advisors and analysts) because they (a) take to long (compared to the political cycle); (b) cost too much; (c) are based on too many unrealistic assumptions; or (d) are too complex or technical. And this problem may be more salient at lower levels of government.

3. Cross-organizational sharing or not sharing (e.g., privacy, confidentiality policies) (or lack thereof) of eco-informatics tools or information. There are two levels of cooperation between organizations: sharing of tools or EI information and co-production of tools or EI information. Collaboration might be more complex in co-production situations compared to sharing situations. Organizations tend to want to avoid paying to develop an eco-informatics tool or information if there is another organization also involved. One group member referred to this as a "tragedy of the commons" problem. And indeed, it is the classic free rider problem in collective action theory.

In addition, organization policies often act as barriers to the co-production of eco-informatics tools or datasets. For example, if there is no recognition in an employee's performance review for effort expended to undertake and maintain cross-agency collaboration in EI tool or information production, the employee may be less interested in undertaking such endeavors.

In some instances, organizations might be interested in sharing information but may be hindered because inadequate metadata have been developed to communicate what the datasets represent. In the context of EI tools such as models, there may a similar interest to share but there may be inadequate documentation (e.g., ontologies) that would promote model sharing and integration with other models. This problem is probably driven in part by a lack of attention in organizations to develop either a carrot or a stick approach for encouraging data owners to produce and maintain the metadata. Examples of "carrot" approaches might be employee performance rewards or positive recognition in Community of Practice situations. "Stick" approaches might be executive order type of mandates with negative consequences to the employee if they are not followed.

4. Communication (or lack thereof) of environmental management decisions grounded upon eco-informatics-based analysis. Organizations sometimes do a poor job communicating pressing issues that have been discovered through EI analysis to decision-makers or the public. This is in part an information diffusion problem regarding the use of the media and other mass communication approaches. Organizations or policy-makers are sometimes caught off-guard by some environmental or ecological problem, and may find themselves facing real difficulties addressing the problem because little or no data exist to help understand and respond to the problem. Or there may be some data available, but the data exhibit a linear trend when in fact the pattern is more complex. A question here is whether there are any EI-related tools and techniques that can help with this kind of situation. These are problems where decision-making or policy is needed to be made under circumstances involving great uncertainty. For a more extensive treatment of this group's work, see https://www.evergreen.edu/bdei/presentations/ summaryPolicygroupfinal.pdf.

2.2 Data Presentation

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. This problem area can be distilled into two major components. One is a model of the role of presentation as the "mediator" between users and their needs and task and data/metadata and their characteristics. The second is the set of research questions and themes that relate to the facilitation of that mediation role.

Essentially, the model suggests that presentation options must reflect dimensions of the user experience as well as the nature of the data but also have their own sets of "constraints" or dimensions that also need to be recognized in presentation. On the User side, presentation types may need to reflect a number of user dimensions: 1) User needs, perhaps conceptualized as tasks, or as time available to user, or his or her context for the activities, and 2) User characteristics, including preferences, dis)abilities, and computing capabilities available. On the Data Side, presentation may need to take advantage of or reflect the nature of the data, the amount of data, metadata available, quality measures associated with the data/metadata, data preparation activities used, policies (such as privacy and confidentiality aspects). Presentation Instantiations and approaches need to reflect the marriage of the user and the data sides. In addition, presentation media add their own "affordances" and issues to the mix that must be recognized. Different software modalities may have different suitabilities for different data types, and different hardware media have different costs, availability, and permanence. These three components to the model will suggest a range of research questions that will help us understand presentation for EI decision making. For more detail about data presentation, see https://www.evergreen.edu/bdei/presentations/wedbo3summary.pdf.

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 out-paces data reliability, and a lack of a network of database professionals upon which resource managers can call for advice or expertise. 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 for them, such as stable standards for data collection and documentation.

2.4 Tools

A major problem that was articulated applies 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) the 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 input data than what it was developed for. 2) the problems of new or different data types and of data collection, 3) the lack of user frameworks and product suites, and development standards, 3) the lack of tools to support metadata issues (creation, quality, etc.), and 4) the 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. See 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. See https://www.evergreen.edu/bdei/presentations/wedbo3HumanCenterednesssummaryc.pdf.

3 Research Issues

Teasing out the research issues from the natural resource management problem space was a four-step process. 1) Participants examined three current research projects to see how they employed interdisciplinary approaches and involved government partners to solve problems similar to those identified above. 2) Workshop participants broke into smaller groups to articulate research issues, which were 3) critiqued by a panel of resource managers and researchers with experience in the *area*. Finally, the smaller groups met to refine and prioritize the issues that were earlier articulated, to identify strategies for sustaining the research, and to find resource management case histories that exemplified the need for the research they identified. For details regarding individual breakout groups, see https://www.evergreen.edu/bdei/presentations/ tuesbreakout2_combined.pdf.

The three NSF Digital Government research case studies were the Forest Portal, UrbanSim, and Understanding Government Statistics (see http://www.cse.ogi.edu/forest/papers/blm-riefing.ppt, and http://www.evergreen.edu/bdei/presentations/borning.pdf and hert_tuesdaylunch1.pdf, respectively). 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 establishing partnerships contribute to data collection, preparation, and assessment, which in turn likely lead to realistic policy scenarios and major policy applications in 2005. The GovStat project models user access to U.S. government statistical information in order to better integrate data across agencies. In building a prototype to harvest government web pages, they emphasized the value of deployed prototypes to identify research challenges, in this case finding data that mapped to user requirements 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 (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. 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.

3.2 Data Quality

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.

The general problem of data quality in decision-making can be summarized as how to determine and communicate uncertainty to decision-makers in studies integrating multiple data sources. 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. This research issues arise in both individual studies and data sharing. For example, in individual studies diverse data sources are combined, and one would like to know the points where error is introduced. 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. Thus, for example: What happens to metadata when multiple sources are integrated? How can metadata management be automated once it is created? How can data standardization help the process of combining metadata from multiple sources? Can open-source tools be developed for mapping data content standards to one another? The research challenge is how general can the tools (that manage data quality in individual and shared studies) 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. By far the most advanced work is being done by the Federal Geographic Data Committee within the USGS, including a biology standard. Metadata standards are well developed and in use by the LTER information managers, and these standards are used in internal reviews of LTER projects. NBII is making a very big push to require metadata using the FGDC standards for its projects. It would be important to pull Valerie Hutchinson into this discussion.

To determine whether uncertainty associated with data synthesis really has an influence 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. This work could be extended by examining how syntheses of data sets are used by decision-makers and how apparent and important the errors were to decision-makers. Specifically, the research question is: 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? An extension of this work could examine how synthetic studies stand up in courts of law in comparison with other forms of "expert testimony."

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? For more detail, see https://www.evergreen.edu/bdei/presentations/wedbo3dataqualityb.pdf.

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; 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 ontologies in ways that go beyond simple versioning.

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. Data collection, ontologization, and modeling embody value judgments; how can computer scientists and developers be sensitized? The semantics of BDE is critical, and includes defining and operationalizing meanings, data standardization, and semantic services. Transferring 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. See https://www.evergreen.edu/bdei/presentations/ wedbo3ontology.pdf.

3.4 Social and Human Aspects

Research issues identified in this area 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).

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. For more detail, see https://www.evergreen.edu/bdei/presentations/wedbo3HumanCenterednesssummaryc.pdf.

4 Conclusions, Broader Impacts, and Recommendations.

One metaphor that could be used to understand the natural resource management vision that workshop participants would like to convey, due to Nancy Tosta, is a fictitious, ideal decision making tool. The tool, dubbed Yoda, helps define decision makers as those who choose among alternatives, and what they do as integrating data –– 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. See https://www.evergreen.edu/bdei/presentations/GreybeardNT.pdf.)

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.

Two critical issues not addressed directly in the workshop, but that emerged as participants followed the agenda set for them by the workshop steering committee are: 1) the issue of feedback loops, and 2) the nature of decision making. One senior scientist urged that if a resource manager becomes more effective at what he or she does, then the effect of the manager on the system he or she manages is not negligible. And, we know very little about this problem.

Understanding the nature of decision making is as critical to those conducting research in this area as the feedback problem. As a result, three of the workshop participants (Bowker, Fulop, and Schweik) worked after the workshop adjourned to develop a primer in decision making for computer scientists and social scientists interested in this research area. 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). A longer report is available on the project web site (www.evergreen.edu/bdei/decisionMakingLong.pdf).

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 team of graduate students at The Evergreen State College observed the workshop, reported on how researchers could articulate educational impacts for their work. 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. The students saw that the ethical issues around large data repositories as a particularly fruitful area for teachable moments. Funding interdisciplinary mission-oriented tasks that force the addressing of local problems was seen as a way to pursue these goals. The students encouraged NSF to partner with agencies that support applied student research. See https://www.evergreen.edu/bdei/ presentations/gradStudentDraft.pdf.

In sum, workshop participants encourage the early focus in eco-informatics and decision making to be on ecological and biodiversity issues, as it was strongly believed that 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 in sustaining and encouraging innovation, research and development in this area. 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?

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.

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. Workshop participants emphasized 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.

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 (NaturServe), William Sonntag, Nancy Tosta (Ross & Assoc Environmental Consulting), Dawn Wright (Oregon State Univ.).

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.

Finally, we acknowledge financial and administrative support from the National Science Foundation (NSF IIS 0505790).