Eco-Informatics and Decision-Making
BDEI Workshop
December 13-15, 2004
Olympia, WA

Human Centeredness
or
Social and Human Aspects of Eco-Informatics and Decision-Making
DRAFT 12/21/04

Advancement of the overall eco-informatics agenda hinges not simply on the advent or adoption of new technologies, but equally on new understandings and formations of the information infrastructures and interrelationships within those infrastructures between individuals, organizations, communities, disciplines, information resources and tools. A new eco-informatics culture needs to be shaped that will bridge personal, technological and organizational gaps among and across various stakeholders, tools and content within the biological, ecosystem and computer sciences; as well as within and across the public, academic and private sectors.

Much of the research conducted under social and human aspects of eco-informatics is expected to rely on the transfer of knowledge from other domains. As an illustration, budgets of telecommunications and consumer electronics companies as well as the defense and the health services industries far exceed eco-informatics budgets. Thus, in many cases the development of new technologies, products, processes and services in the biodiversity and ecosystem community may depend on transferring and adapting concepts from these larger, more established enterprises.

Identification of best practices within and outside the BDEI domain is also recommended as a useful knowledge capturing and transfer process. Some individuals and organizations have, over time, achieved a significant degree of successful information management practices, often through trial and error. These stories of best practices are highly regarded by the user community as many are replicable for their situations.

Refining the Problem Space:
Note is taken that social and human aspects of eco-informatics related to decision-making processes are numerous and frequently broad in scope. The following research issues are offered as a sample of the potential areas available to those wishing to engage in research regarding the social and human aspects of eco-informatics.

Research Issues and Questions:
Collaboration and Information Sharing
Numerous collaborative efforts achieve successful results, but little is known of the factors that contribute to that success. Some research questions that need exploration include:

- What needs to be in place to enable or facilitate successful collaborative efforts?
- How do we measure success?
- What are drivers, motivators and incentives for organizations and individuals to engage in collaboration?
- Are there effective institutional designs for interpersonal or inter-organizational sharing and collaboration?

- What kinds of models are out there?
- What key variables lead to success?
- What are the potential disincentives (e.g. costs, privacy, property rights, agency branding) to collaboration and how are they dealt with?

Models for successful collaboration exist and need to be analyzed, documented and transferred to the BDEI domain. Examples include; the Open Source programming domain, the Internet, meteorological information sharing, and the National Bird Count.

Conversely, it is also recognized that many collaborative attempts end in failure. Yet, little is documented as to why these efforts have not been successful. Identifying obstacles to collaboration be they social, political, personal, or perhaps technical, needs to be better understood by members of the BDEI community.

Human-Computer Interaction
Eco-Informatics tools need to have very smooth human use interfaces in order for humans to adopt them. There is currently little connection between informatics tool developers and the scientists and decision-makers that require those tools. This gap needs to be closed.

- Can we take the tools currently available, put them together and make it possible for discovery and use in a way that is relatively simple?
- What can be learned or adopted from metadata management, standardization, exchange mechanisms and knowledge management systems (the level of criticality for the user, basic screening, deploying studies to making public policy decisions, etc.)?

Research opportunities exist along the “chain of custody’ of information from project planning to data collection (field measurement and collection models and tools) to information presentation (visualization toolkits) and decision-making. Other IM steps that require improved HCI include metadata authoring and management, data analysis, quality assurance, predictive modeling, archiving, information access, knowledge management, and requirements for locating and utilizing the nation’s vast and disparate repository of biological legacy data.

Incentives and Regulations
Research has taken place on what effect incentives or regulations play in achieving business and management goals. Yet little is known within the BDEI domain of the effectiveness of incentives and regulations on information management practices.

- Can traditional structures, e.g. “publish or perish”, in the eco-sciences facilitate information management?
- What role do organizational policies, regulations or procedural guidelines play and how effective are they under varying conditions?
- How do incentives and regulations cut through cultural resistance to sharing information or participating in an IM agenda?

Each organization and management unit may adopt different approaches with differing degrees of success. Research issues include identifying best practices that could be replicated and tested in BDEI organizations.

Education and Training

Research under this domain includes broader educational aspects such as how to broadcast the value of sound data management within and outside the BDEI domain and the organizations therein. Failed attempts to institute data management procedures and programs sometimes rest on the lack of understanding of the value and benefits of data and metadata. How best to infuse data stewardship values into organizations and individuals is a key question. And further, what educational approaches seem to work for the different levels of stakeholders in BDEI.

Biologists and other data owners are frequently educated in the process of collection and analyzing data yet are not conversant in cataloguing, archiving, dissemination, managing and other aspects of sound information management. What extended learning opportunities should be developed? What would they encompass? Do they include "communities of interest" groups or "centers of excellence"? Do models or best practices for such educational opportunities exist? What approaches are best for data management training on eco-informatics tools?

This category also includes promotional and outreach efforts. Information management requires that all stakeholders are aware of, and "buy into" the value of sound information management. These stakeholders include mid- and upper-level managers and support staff that may not necessarily have a direct involvement in IM activities, but may develop budgets, sign directives and policies, promote information availability, approve training seminars, etc. This group includes administrators, resource managers, budget assistants, and others within the "resource management organization". How best are these stakeholders kept knowledgeable of informatics benefits and requirements? How do potential solutions resonate with BDEI stakeholders? How is "brand" confidence built into BDEI programs and information systems?

User-Needs

All too often, information systems and programs are developed without the input from the intended users. A large percentage of failed information systems did not incorporate user requirements in their design. What specifically is required in user studies to ensure long-lived and usable information management programs? How detailed do user studies need to be? Are any processes or models available to BDEI that facilitate better communication and coordination between users, system developers, information managers and information providers? A significant number of studies in other domains have been conducted. Are adequate models and tools for user requirements and usability testing available? How can such models and tools be adopted, modified and tested in the BDEI domain?

Researchers in this field need to consider the users required steps for successful decision-making including; identification of appropriate sources, evaluation of sources for relevance and reliability, extraction of required information, manipulation using appropriate tools, interpretation of results and presentation to appropriate audiences.

Scenario #1 – A State Agency Official Trying to Prioritize Parcels for Conservation: Modeling the Development Fringe

Jane Doe is an analyst for a state agency (or a conservation NGO) who is interested in forecasting landuse change over a region with the hope that they can identify habitat parcels that lay on the “development fringe” and are most threatened by human encroachment. Jane also wants to be able to explore how changes in particular state policies might influence future development patterns.

Jane knows a landuse change model might help in this direction but would rather not build one from scratch and knows that there is not enough expertise in her particular organization to begin such an endeavor anyhow. Jane decides to look for existing, off-the-shelf models that might help answer her question.

Jane does a review of existing landuse change models and finds some good news --- there are quite a few landuse change models already in existence. These models vary in the questions they were developed to answer, the theoretical underpinnings, the geographic and temporal scale that they operate, the data inputs they require, and their modeling approaches and technologies. Some are offered by private firms as a proprietary package and others are licensed by their developers as Free/Libre or Open Source (FL/OSS). Some are a hybrid, requiring a proprietary software platform but the model itself is available through a FL/OSS license.

Jane’s review reveals a couple other issues. First, many of the available models appear to be missing some of the key “policy related variables” that she would like to explore or change in various policy scenarios. Some use proxy variables to capture human decision-making behavior (e.g., demographic information) but these variables are not really ones that can be readily changed through public policy. Second, in reviewing literature on various models, it appears that in some instances, there has been limited or no growth in the deployment and use of the model beyond the original group who developed it. Jane expects that part of that reason is because of the significant transaction costs it might take to learn how to use the model, the kinds of disciplinary expertise required, the technologies and approaches used, and the cost.

After some thought, Jane decides the transaction costs to possibly purchase, learn and apply a landuse change model to her interest area are too high, and gives up on the idea.

Possible collaborative scenario of the future

Fast forward ten years. Jane is in the same position (she likes her job!) and decides to revisit her idea again. She initiates another survey of the landuse modeling environment and finds a very different situation.

She does an Internet search, and discovers that a collaborative infrastructure or “commons” exists that is devoted to the production of landuse change models. Components of this infrastructure include:

- A library of existing models (and sub-modules), some proprietary and some licensed FL/OSS, along with metadata on those models and ontologies that describe model structure. This is a kind of “market of models” with some metrics on how many people are using them, etc., that give her some idea of their utility.

- A library of theoretical and empirical papers about these models or model components. In fact, this looks a lot like a refereed online scientific journal.
- A library of data that people have been willing to make open access.
- A library distance learning material and services. Here she discovers material that is open access and licensed as “open content” but she also discovers private companies who are providing services to support some of the learning or application of these models.
- A “collaborative development” system for each of particular models. This appears to follow many of the principles of “content management systems” (e.g., version control, bug reporting, etc.) but has features like this not only for model modules, but also for other model project components, such as theoretical papers or distance learning materials.
- All of these components appear to have gone through some formal peer-review process in order to be posted on the repository. In other words, to Jane, this looks a lot like an online journal except that much more than final papers are published. Jane thinks that one reason for this is that this “publishing” of various modeling-related products might provide an important incentive for scientists to contribute.
- Importantly, all of the products that are stored in this commons have associated metadata that associates how a component (e.g., model module, theoretical paper, dataset, etc.) are licensed. That is, rules for use and further derivations are attached. Some products might fall under a GPL license (free as in freedom software) that promote free distribution and new derivatives but require the derivatives to be licensed the same way. Others might fall under a less restrictive license (in terms of its ability to be used in a commercial package). Interestingly to Jane, some Creative Commons licenses are used for theoretical and other papers about various models, and Jane discovers that some people have taken a paper, and made a new derivative of that paper that better fits a particular empirical setting. For example, she finds a paper on the drivers of landuse change in the western U.S. and another on a Brazilian context where the fundamental paper is the same but the theoretical drivers are different.

  The various library systems for various components (models, papers, distance learning, data) keeps good documentation on how these versions have changed and who contributed the intellectual property and Jane notices there are new metrics in the system on how people have contributed intellectually and instead of papers being cited sometimes it is the logic of a model module or “snippet” within a second derivative paper. Empirical papers and most data, however, appear not to follow this logic.

- Jane notices that it is not just programmers or modelers collaborating in this “ecoinformatics commons”. But rather, there are modelers, scientists (theory), data providers, “citizen scientists”, etc.
- Jane also notices for each model, there is some governance structure making decisions about what goes into the next version of the model, developing standards or norms of behavior in terms of development, resolving conflicts, etc.

- Jane notices that in some model cases, Government agencies or companies have contributed significantly to their development, giving the impression that they are supporting this commons by actually paying some of their employees to contribute to further development of models held in this repository.

- Jane also notices that the commons infrastructure has some sponsorship, but it isn't clear how that is supported (financially).

What this collaborative system allows Jane to do is find a model that best suits her needs. She might follow the distance learning material first, or pay for some services to help her apply the model. She might read theoretical papers and other documentation about the model to make the decision.

Reading the documentation on building the model application she finds there are some places that need clarifying, and she decides to modify a document to improve it and because it is licensed as "derivative works ok" she downloads it, modifies it (documenting where her modifications are) and submits it back to the community as a new version based on what she has learned. She is now contributing back to the commons.

When Jane runs into a problem she can ask the community using the model for help or turn to the commercial service providers.

She also decides in her case that a logic to include a particular policy variable is not currently in the model, and places the request on the collaborative system. Someone else sees this request and implements it.

Jane eventually documents her experience with the model in her application – she writes an empirical paper about her particular application and submits it for peer-review.

The result is that Jane has a model that satisfies her needs. In doing so, she has also contributed to its further growth and development. Her use of a particular model has added an additional person or group to the community who is interested in that model building its ranking in the "market of model" hierarchy.

What research is required to make this work?

- Research into the development of next generation collaboration tools. There are open source components of this already existing… e.g.,
    - Plone/Zope content management
    - Library management systems
    - Software content management (e.g., CVS)
    - Open source E-journal software (e.g., open journal? Used by Ecology and Society)
    - Wikis
    - Creative commons licenses
    - Nothing to my knowledge that would support the whole system, including peer-review of models, data, papers, etc.

Scenario #2 - Death Valley National Park Devils Hole Pupfish - Resource managers requirement to take disparate legacy data and explore its utility for making modern management decisions.

The world's entire population of Devils Hole pupfish (*Cyprinodon diabolis* Wales) lives in a unique groundwater filled limestone cavern in Death Valley National Park. Beginning in 1968 large-scale agricultural pumping drew down the water level in Devils Hole, threatening the fish with extinction. The National Park Service and local advocacy groups took the fight for the fish to the U. S. Supreme Court. In 1976 the court ruled for the pupfish and ordered the protection of their groundwater resources.

Recently the pupfish population has again declined dramatically; the cause of the decline is unknown. The population has steadily deteriorated from a high of 541 fish in 1994 to 219 fish in 2004. In addition, on September 11^{th}, 2004 a flash flood and heavy debris flow caused high fish mortality and substantial habitat alteration. The changes in the pupfish habitat may prevent successful recruitment during 2005, which may lead to the extinction of these short-lived, endangered, fish.

Despite thirty-five years of data collection and research regarding Devils Hole and its resident pupfish, the National Park Service lacks the information to answer such basic questions as: What is causing the recent decline in fish numbers? Should the NPS manipulate the habitat to replicate pre-flood conditions? What techniques will allow aquarists to successfully breed these fish in aquaria? Without such information the National Park Service risks making incorrect management decisions regarding the fish. Information that could aid in decision-making may be embedded in existing data and records, but at present this information is not easily locatable or, if available, it is not in formats usable for analysis and decision-making. Death Valley NP seeks assistance in identifying and developing appropriate informatics tools to facilitate discovery, organization, synthesis, and preservation of information that would benefit the protection of the Devils Hole pupfish.

Informatics Issues:

1. A need for integration of data over time (65 years) and across agencies. Agencies maintain their own standards, tools, protocols, scales, formats, archives, etc. for their databases.
2. Modeling and simulation tools are required to determine, among other things, whether the current accumulation of sediment is a result of monitoring efforts by humans or through natural causes.
3. There is a need for the ability to accurately measure the quality of legacy data prior to be utilized in decision-making processes by the multi-agency Recovery Team.
4. A need for visualization tools to show the project the shelf and recruitment rates under various management scenarios.
5. A way to look at similar studies and possibly identify data that should be collected, but has not, that might help the Recovery Team gain a better understanding of why numbers are declining.
6. Previously, population decline was predicted by declines in water levels. Water levels have stabilized, but population continues to decline. Need to identify and test better indicators for projecting population declines.