Vision statement on a National Ecological Data Center

Note: This statement was developed at the request of the ESA Governing Board by a subcommittee consisting of Grimm, Palmer, & Inouye. Much of this material is from the ESA Visions Report and thus has been vetted by many members. O.J. Reichman drafted or edited most sections of this part of the Visions report and provided comment on a draft of this statement.

Ecological research questions extend across a spectrum that ranges from increasingly rich and sophisticated molecular-level analyses of both living and nonliving components of the world's ecosystems to integrated views of the entire globe. Ecological analyses, and management decisions that rely on them, commonly employ highly diverse data. Unfortunately, much of this information has been accumulated and stored in haphazard and inaccessible forms in the nooks and crannies of research labs. This may not have been recognized as a big problem before the arrival of the digital age, but now it seems an archaic barrier to progress. Even when made available, environmental data are often highly distributed and profoundly heterogeneous, and few tools exist that can acquire and characterize data and models, and then make them widely accessible in a convenient, integrated way. Much of the information that already has been gathered is no longer in useful formats, even for those who collected it. These barriers to data access limit progress in environmental science, and greatly limit data use by resource managers, economists, political analysts, and other decision makers.

Paralleling and perhaps more challenging than the problem of access to existing data is the coming deluge of ecological information that will accompany the explosion in new technologies for observing ecological systems. Both the establishment of coordinated observatory networks, such as NEON, and expanded capacities for individuals and institutions to gather spatially and temporally detailed data, presage an urgent need to document, store, retrieve, manipulate, analyze, and display data. ***In the future, a community of ecologists will freely share information and possess the tools to access and use this information to expand conceptual frameworks, build models, and construct scenarios that can help decision makers envision the consequences of human actions.*** As described in the Visions Report, building this community capacity will require establishment of a number of centers that foster collaboration and coordination, efficient use of extensive ecological data, possibilities for education, and knowledge exchange within and outside of the scientific community.

A revolution in information technology that will ease the generation of ecological knowledge is occurring in response to this growing need to manage ecological information. Information technology research is yielding mechanisms for ecologists to document their data using standardized metadata[1] protocols, and store their data in carefully managed, widely accessible archives. These data can then feed into the conceptual frameworks and analytical tools described above. Thus, ecology could greatly profit from the development of technological means, cultural inducements, and training opportunities to represent ecological knowledge effectively. This can be done via a suite of widely available tools for the seamless integration of data, theory, concepts, and models. A National Data Center is a key component of this vision.

We here reiterate, update, and expand upon several steps toward the implementation of ESA's vision of data discovery and access.

Approve the proposal to establish a data registry. A proposal for the ESA Data Registry is under consideration by the Governing Board. The first step to a broader vision for ecological data

[1] Metadata is information about data; see Michener et al. 1991

discovery is to begin to change the culture through a data registry. The proposal stipulates a requirement for data registration for all papers submitted after January 1, 2006.

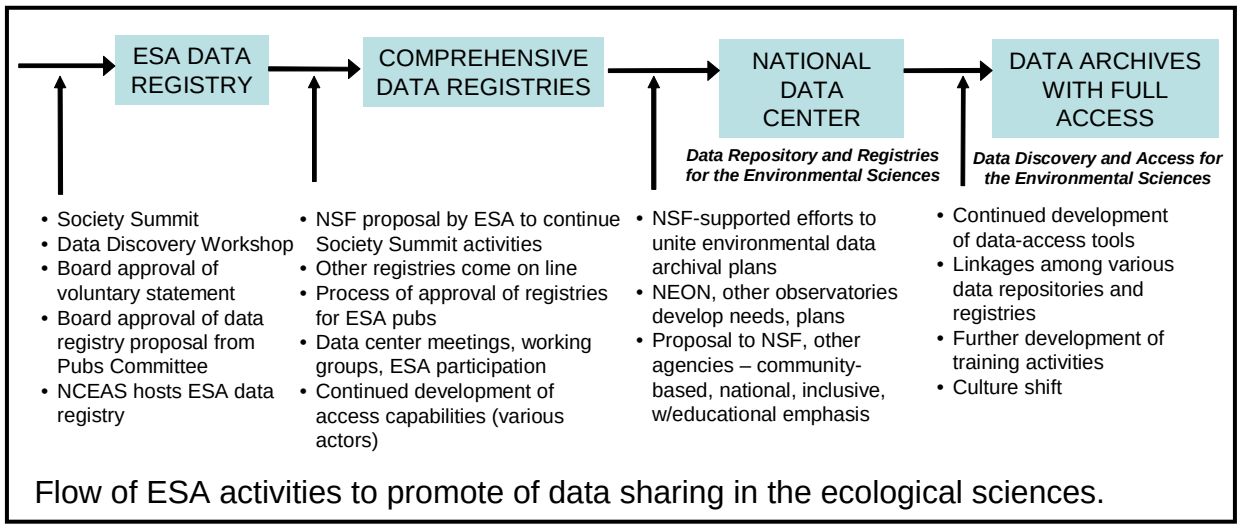
Make raw data and metadata easily and freely available. The ESA should set a goal that all papers published in its own journals include links to metadata (i.e., are registered) by 2006. By 2008, all papers should make associated raw data and metadata easily and freely available.

Lead an effort to encourage federal and state agencies, NGOs, research centers and field laboratories and their scientific societies to follow suit. ESA is already so engaged via its participation in the Society Summit and leadership in a follow-on proposal to continue these discussions in a Joint Working Group on Data Sharing. These efforts should be continued.

Encourage training in ecoinformatics both during graduate school and in ongoing programs. The ESA should promote ongoing training through regular training workshops and opportunities at its annual meeting. Several informatics research projects (e.g., SEEK, KNB, RDIFS) have created training opportunities at ESA, but these are typically lightly attended. Two such workshops were cancelled in the Montreal 2005 meeting due to insufficient interest. So in addition to providing the outreach opportunities, we need to establish ways to motivate ecologists to participate.Workshops might be a component of any externally funded data sharing projects initiated by the ESA.

Work toward the establishment of a National Ecological Data Center. Such a center would be the repository for metadata and data sets, a creator of data catalogs, and a clearing house for information on data availability. It would be a center for technology development in ecological cyberinfrastucture and a hub for training in ecoinformatics. Finally, through development of tools for rapid data access, analysis, and visualization, the NEDC would interface with other exisiting and planned centers that collectively enhance community capacity. Clearly, the magnitude of this effort will require cooperation and concerted effort across a broad segment of the ecological research community, including close interactions with NCEAS, the LTER network, NEON, other scientific societies, and funding sources (especially the NSF). An educational emphasis of the NEDC should extend to the undergraduate level.

Broaden access and discovery activities. Ultimately, the national ecological data center will serve as one of many focal points for community sharing of information, development of models, and exchange of information between ecologists, scientists in other disciplines, policymakers, and the public. Sophisticated means of linking and accessing information will continue to be developed. Training in ecoinformatics will become standard in the educational trajectory of all ecologists, effecting a culture shift in the discipline.



Flow of ESA activities to promote of data sharing in the ecological sciences.

The rationale for a National Ecological Data Center

Data repository. Problems of access to extant data, coupled with the need to prepare for an expected flood of data from new environmental sensor networks and other technologies, demand a coherent and community-based plan for their solution. The NEDC would provide a centralized, community-based location to both ensure the long-term integrity of data stored and provide access with the most relevant and state-of-the-art tools. Distributed models also are possible; for example, the NEDC could expand upon and link existing repositories such as the ESA's Data Archives (repository for published and peer-reviewed data), the NCEAS data repository, or the LTER network database, among others.

Technology development in ecological cyberinfrastructure (ECI). ECI provides the link between people and the data, hardware, software, and other tools for creating and displaying ecological knowledge. Advances in ECI require the development of generic tools for data input, access, and analysis. Ecological data should be digitally captured early in the acquisition process, while concurrently generating metadata. Data then must be easily and widely accessible to individuals even in the most remote areas of the world through a set of generic data access tools. The goal is to link the vast body of data resources, whether they are small compilations or large, well-known data sources. The key approaches must include an efficient, flexible, and standardized way to describe ecological information, a powerful information searching capability, and an integrated process of information analysis and modeling that completes the stream from raw data to processed data and publication output. One important result will be the enormous savings in the ecologist's time—time that then could be devoted to more forward-thinking research.

Ecoinformatics training. The creation of standardized languages for the Internet, for example XML, has allowed the development of specifications that will facilitate information coding and retrieval by search engines and interoperability of software tools. Indeed, such standards have beendeveloped, such as Ecological Metadata Language (EML). However, most ecologists are unfamiliar with these tools and techniques. To manage our data correctly and to learn from the data most effectively, ecologists will require easy access to user-friendly software, a publishing and funding system that requires or rewards such behavior, financial support for information management as part of any research funding, and the development of secure data archives. Many of these elements can be incorporated into a coordinated training plan that has the objective of maximizing ecologists' use of ecological data.

Tools to interface with other community-based efforts. Once scientists acquire datasets, they must be able to analyze them quickly, and transport them into appropriate visualization tools that make the data understandable. It will then be possible to find and download data rapidly, and analyze disparate datasets to test an idea much in the way we now use abstracts of articles to get a sense of whether a new idea warrants our further consideration. Sophisticated visualization tools, including maps and landscape-based Geographical Information System (GIS) data and virtual imaging, will make communication within and beyond the scientific community easier and more effective, and allow major advances in ecological modeling. Visualization is particularly useful in making ecological knowledge available and ensuring its consideration by political entities and decision makers. We expect that many of these steps in the use of datasets to advance conceptual development, test specific hypotheses, or exchange knowledge with decision-makers, will take place via a close interaction of the NEDC and other entities such as the existing National Center for Ecological Analysis and Synthesis or a proposed Center for the Ecological Implementation of Solutions.

Through improved generic data input, access, and analysis tools, standardized metadata, and open access to environmental data, more comprehensive analysis and synthesis of ecological knowledge will be possible. The beneficiaries will be both the original gatherer of the data and the science world at large, but also a global society that increasingly seeks sustainable solutions to environmental problems.