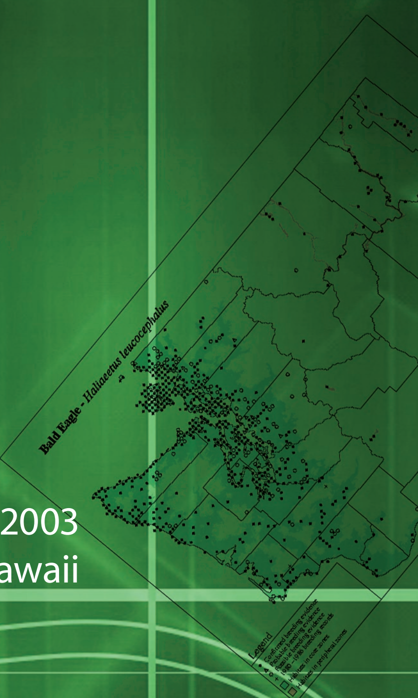
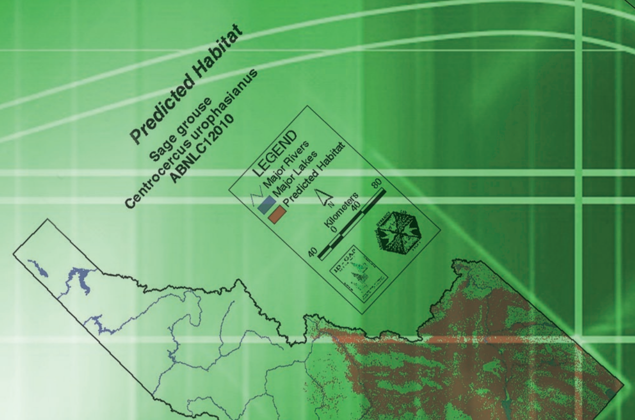
First NBII Biodiversity Modeling Workshop

Results and Recommendations

Proceedings of a workshop held July 27-31, 2003
Kihei, Maui, Hawaii





First NBII Biodiversity Modeling Workshop

Results and Recommendations

Proceedings of a workshop held July 27-31, 2003
Kihei, Maui, Hawaii

Sponsored by
The National Biological Information Infrastructure (NBII)

Convened by
The U.S. Geological Survey Center for Biological Informatics (CBI)

Hosted by
The NBII Pacific Basin Information Node (PBIN)

Editors:
Bradley O. Parks, Director, NAC-BDEI, University of Colorado / Argonne National Laboratory
Mark D. Fornwall, Director, NBII Pacific Basin Information Node, U.S. Geological Survey
James F. Quinn, Director, Information Center for Environment, University of California-Davis and NBII California Information Node

Please cite this publication as follows:
Parks, Bradley O., Mark D. Fornwall, and James F. Quinn, eds. *First NBII Biodiversity Modeling Workshop: Results and Recommendations.* Proceedings of NBII Biodiversity Modeling Workshop, July 27-31, 2003, Maui, HI. Prepared by the North American Consortium for Biodiversity and Ecosystem Informatics (NAC-BDEI) at the University of Colorado at Boulder for the National Biological Information Infrastructure. Denver, CO: U. S. Geological Survey, Center for Biological Informatics, 2004.

Production and editorial assistance from the publication staff of the NBII National Program Office.

Table of Contents

Foreword . . . v

I. Introduction . . . 1
- A. Problem statement – biodiversity modeling needs . . . 1
- B. Position statement – NBII investments and interests . . . 1
- C. Workshop description – terms of reference . . . 1

II. Findings . . . 3
- A. Modeling biodiversity – status and expectations . . . 3
  - Recommendations
  - 1. Implement a clearinghouse for existing biodiversity models . . . 4
    - Proposed constituent actions:
    - a. Build classifications of critical ecological (functional) attributes of species . . . 4
    - b. Establish a decision support system for selecting and using biodiversity models . . . 6
    - c. Track reuse of biodiversity models . . . 8
    - d. Enable and promote peer review of biodiversity models . . . 9
- B. Data and information for biodiversity modeling - status and expectations . . . 9
  - Recommendations
  - 2. Identify and catalog the relevant data universe . . . 9
    - Constituent actions:
    - a. Automate competent data discovery . . . 10
    - b. Assess information needs and data gaps . . . 10
    - c. Advocate and fund data digitization . . . 11
  - 3. Improve data integration through schema element or ontology reuse . . . 11
    - Proposed constituent actions:
    - a. Develop and manage a federated data schema . . . 11
    - b. Registry support and data services . . . 13
  - 4. Promote data quality, data exchange, and peer review . . . 14
  - 5. Provide education on biodiversity information management and modeling . . . 14
    - Proposed constituent actions:
    - a. Use metadata to educate model users . . . 14
    - b. Develop multi-scale consciousness and ecological perspective . . . 14
    - c. Educate collaborators about biological and ecological informatics . . . 15
    - d. Train and expose users to biological and ecological informatics . . . 15
    - e. Develop a teaching text . . . 15
    - f. Found and sponsor an electronic journal . . . 15
  - 6. Provide a service outlet to develop models and maps operating through distributed biodiversity modeling systems . . . 15
    - Proposed constituent actions:
    - a. Provide scaled, remote computation service . . . 15
    - b. Provide mapping and modeling services via the Web . . . 16
    - c. Build user interfaces using interoperable middleware . . . 16
    - d. Implement distributed biodiversity modeling systems . . . 16

III. Implementation 17
A. Adopt terms of engagement and proposed NBII Action Plan - status and expectations 17
Recommendations:
7. Adopt a leadership role in improving biodiversity modeling 17
8. Proactively and systematically engage the biodiversity community 17
9. Adapt NBII institutional interfaces to ensure successful leadership 17
a. Designate leaders/managers/and ombudsmen 17
b. Make NBII's contact persons apparent and accessible 17
c. Define responsibilities and chain of authority. 17
d. Orient staff and management with ongoing internal education efforts 17
e. Establish, monitor, and implement strategic and tactical plans 17
10. Begin three-year start-up phase with proposed schedule of effort. 17
B. Concluding Remarks 20

IV. Appendices 21
A. Acronyms used. 21
B. Workshop participants 22
1. Contributors in attendance
2. Contributors not in attendance
C. Workshop prospectus 22
D. Workshop program. 23

Foreword

The National Biological Information Infrastructure (NBII) is being implemented to serve national biodiversity management, research, and education needs. It began in the middle 1990's as a World Wide Web-based information system developed through partnerships with key organizations. The U.S. Geological Survey (USGS) Center for Biological Informatics (CBI) serves as an infrastructure node and provides leadership for NBII implementation. The NBII seeks to provide a means to access the vast amounts of scientifically credible, biological resources data and information residing in virtually thousands of organizations ranging from federal, state, and local governments to universities, museums, and many non-governmental organizations.

Over the past ten years the NBII has made considerable progress toward the goal as originally envisioned. In order to be responsive to needs of various users, the NBII has developed in accordance with the recommendations of the National Research Council.[1] Guidance has also been provided by the Presidents Committee of Advisers on Science and Technology.[2] Additional recommendations have emerged through collaboration with other national (BioEco, Biodiversity and Ecosystems Informatics Workgroup of the Committee on Environment and Natural Resources, National Science and Technology Council) and international bioinformatics organizations (Global Biodiversity Information Facility [GBIF] and the United Nations Convention on Biological Diversity [CBD] Clearing-House Mechanism). The result is a distributed national information system built through regional and thematic focal areas (referred to as nodes). This approach allows the system to simultaneously address both geographic and issue/species-based biodiversity needs. However, as stated in the documents referred to above, the NBII must continue to evolve into an integral component of the biodiversity enterprise.

Modeling biodiversity is a rapidly advancing method used to further our basic understanding of both species distributions and ecosystem functions. The discipline is data dependent and builds on our current knowledge through experimental hypothesis testing or by forecasting possible changes in landscapes and habitat as a result of human activity (both intentional and unintentional). As interest in these methods expands, there will be increasing need for improved data access and better biodiversity models to support both basic research and applied biodiversity management. Toward this end, the NBII has chosen to embark on a series of workshops to define its role and contributions in support of biodiversity modeling, a logical next step in its mission to support biodiversity enterprises.

Specifically this workshop seeks to address two NBII objectives:

- Developing a suitable framework to support knowledge discovery and creation for the nation's biological and ecological resources; and
- Leading the development, selection, and distribution of tools and standards necessary to facilitate interoperability and allow meaningful interactions with scientific data and information.

Both objectives require the NBII to provide access to key data sets for applications and to enable development and distribution of models and tools to apply those data to address needs of biodiversity enterprises.

This workshop is the first of several planned by the NBII to investigate the range of possibilities in biodiversity modeling thought, practice, and collaboration. This first workshop was intended to lay the groundwork and we feel that mission was accomplished. It lays out a very broad vision, one that will require considerable advancement in both biodiversity modeling and informatics before it can be fully realized. Additional work is required to define a clear road map with discrete, attainable goals that will lead us toward our initial core vision. Early efforts will go into products showing proof of concept for distribution modeling as an attainable, highly visible, well-defined goal for the whole biodiversity informatics enterprise.

These are ambitious recommendations – the beginnings of a grand vision – that reach beyond the NBII, as currently constituted, and probably beyond currently available resources. In practice, these needs and recommendations are global ones, and progress will depend upon collaboration and priority-setting among a plethora

---

[1]National Research Council, *A Biological Survey for the Nation* (Washington D.C.: National Academy Press, 1993).

[2]President's Committee of Advisers on Science and Technology (PCAST) Panel on Biodiversity and Ecosystems, *Teaming With Life: Investing in Science to Understand and Use America's Living Capital* (Washington, D.C.: Executive Office of the President of the United States, 1998).

of groups facilitating standardization and sharing of biodiversity data, including the GBIF, multiple initiatives in the museum community, a variety of international government-to-government efforts to achieve the goals of the CBD and subsequent treaties (for instance, the Inter-American Biodiversity Network [IABIN]), and some specific efforts focused on invasive species (Global Invasive Species Program [GISP] and various bilateral efforts).

It is also recognized that all viewpoints about biodiversity modeling were not included in this first workshop. This is partly a result of our initial decision to constrain ourselves to the issue of modeling species distributions, rather than considering much broader issues of “forecasting” for all of biodiversity and ecosystem informatics (BDEI). Within this narrower domain, and subject to time and resources available, we have tried to include as many view points as possible, both in the workshop and in subsequent review. We hope an inclusive approach is evident in this report and its “positioning” as only the opening discussion in what may be a series of workshops on this topic. The workshop was also, inevitably, limited by the availability of key participants. Due to time and travel considerations, some participants were unable to attend. One consequence of this, to name one compelling example, was that considerable attention was devoted to federated schemas, limiting what might have been said about ontology-related approaches (both approaches are evolving and we do not endorse either approach at the expense of the other).

Readers should note that, in the spirit of opening an inclusive discussion about biodiversity modeling, review comments by contributors have been preserved in the narrative of this report in order to maintain a transparent focus on important issues and productive differences of opinion, rather than pressing for premature closure. And, in that sense, even the recommendations (NBII recommendations noted in green text) provided in this document are intended as means, not ends.

This First Biodiversity Modeling Workshop was an important step for the NBII, one that depended upon support by those who attended workshop. We wish to thank all workshop participants and contributors (see Appendix B) for their valuable involvement with the workshop and their investment in the future direction of the NBII.

I. Introduction

A. Problem statement – biodiversity modeling needs

A recurring challenge to biodiversity analyses is that known occurrences of ecologically important biological elements (habitat types, species, or genetic types) typically represent only a fraction of the geographic range in which they are actually found. Furthermore, those known occurrences may also reflect substantial sampling artifacts, and they may be dynamic in space and time.

In most cases, inferring the full present and future distribution of these elements is impractical or infeasible from field sampling alone, and must rely on a potentially complex set of modeling, mapping, visualization, and other tools. Important applications of tools for determining species distributions include setting geographic priorities for protecting rare species, genotypes, and habitats; identifying sites at particular risk for the establishment or spread of invasive species; aiding the discovery of new populations or genotypes of high conservation value; and inferring the impacts of future changes in environmental policy, land use, or climate.

Within a community representing overlapping interests of biological resource information providers (NBII, BDEI, CBI, World Data Center [WDC], National Invasive Species Council [NISC], and so forth), there has been some conceptual convergence in approaches to this class of analyses. However, these have not coalesced into a coordinated research agenda, despite some pilot collaborations. The purpose of convening a workshop was to assemble researchers and practitioners in biodiversity modeling, both to design such an agenda, and to identify and launch high priority demonstration or start-up projects. Following are suggested points of discussion and problems/issues of common interest:

- Review models in general use, or under construction, for specific application to species distribution.
- Consider infrastructural actions that the NBII could undertake to aid adaptation or implementation of these modeling efforts.
- Identify and rate opportunities for useful support and collaboration.
- Identify and rate training, exchange, and possibly intern, postdoctoral, or sabbatical needs and opportunities.
- Identify and rate funding needs and opportunities.
- Recommend processes by which the NBII might jumpstart distributional modeling as a core initiative.

B. Position statement – NBII investments and interests

As a national biodiversity information system, the NBII supports biodiversity science enterprises by both providing access to and enabling application of biodiversity data and information. At the same time the planet and its biota and ecosystems continue to rapidly change. This rapid change has put added pressure on scientists and resource managers to improve their knowledge of the current distribution of species, as well as to be able to forecast possible future changes in species distributions and the impacts of those changes on ecosystem functions. Thus, modeling of biodiversity is increasingly being used at local to national levels for conservation research and decision-making. With the complexity of questions confronting conservation communities and the early stage of this phase of the NBII's evolution, the time has come to contemplate the ability of the NBII to support efforts of the biodiversity modeling community.

The NBII has been directly involved in ongoing modeling efforts associated with state and regional gap analysis activities under the Gap Analysis Program (GAP) and related cooperative projects. Support to the broader biodiversity modeling community is a logical outgrowth of these efforts. Such involvement may include many elements, from community development, education and ad hoc meetings, through data needs assessment and standards promotion and development, to the possibility of supporting significant computational infrastructure. As an initial assessment of the NBII's potential future involvement in biodiversity modeling, this report:

- Elaborates the range of problems challenging contemporary biodiversity modeling, providing a framework for formulating the responses possible by the NBII; and

- Provides a first draft comprehensive strategy and schedule of efforts, as an aid to planning logistics, resource needs, priorities, and implementation.

C. Workshop description – terms of reference

A working meeting was convened by the CBI from July 27 – 31, 2003, hosted by the NBII Pacific Basin Information Node (PBIN) in Maui, Hawaii. The meeting assembled 21 persons (2 others could not attend), representing many agencies and interests, to evaluate the status and prospects for biodiversity modeling. This document reports the conclusions of participants and contributors. It is expected that participants in subsequent workshops will assess, modify, and build upon these findings using them as a point of departure, baseline, and early organizational framework.

Participants were selected to provide the broadest possible representation in terms of professional and scientific diversity of viewpoint, geographical and institutional variety and inclusiveness, subject to the maximum capacity of facilities, and budgetary, travel, and scheduling constraints. The meeting was designed as a forum for collaboration among qualified scientists and technologists specializing in modeling biological phenomena, taxonomic classification, biological data management (biological informatics), spatial data analysis (primarily using Geographic Information Systems [GIS]), information technology applied to biological informatics, and practical biodiversity preservation and conservation. Despite constraints on participation, basic criteria were met and distribution among these specialties was maintained. Plans to convene periodic workshops to continue to examine biodiversity modeling will allow further optimization by incrementally defeating these constraints.

The working protocol for the meeting emphasized active discussion, reconciling competing viewpoints, and collaborative writing to prepare these consensus recommendations. The program defined a progressive, intensive, and goal-oriented meeting. All participants were required to work from objectives (see Problem Statement above) to agreement about useful outcomes (see Recommendations following), and priorities for clear action (see Schedule of Effort). Those hosting and those contributing deemed the meeting necessary and successful, demonstrated by the long hours worked. Participants expressed strong hope that the recommendations provided here will be adopted and/or adapted for implementation by the NBII and its institutional, policy, science, and management partners.

The following recommendations have been added to capture maximum benefit from the workshop, to provide for effective implementation, and to bridge to subsequent meetings. These have been prepared by the core group of meeting organizers and ratified by other meeting contributors and participants.

Those agencies and individuals accepting responsibility for the actions recommended in this report should:

- Actively solicit review of this document by additional contributors.
- Publish and disseminate findings (in whole or in part) in varied scientific and lay communications such as journals, newsletters, and press releases.
- Open comment mechanisms enabled via the Web, to assist dialog between workshops/meetings.
- Retain continuity by inviting a core group of participants to collaborate in future/subsequent meetings.
- Document deviating, competing, and alternative approaches and decisions as they accumulate.

II. Findings

A. Modeling biodiversity - status and expectations

Need to understand the distribution of taxa

Species, community, and biome distributions are in a constant state of flux. Understanding this persistent change (particularly man-induced versus natural variations) at a variety of spatial and temporal scales is problematic but essential to our understanding and management of biodiversity. Our knowledge of what-lives-where is still incomplete and limited largely to sparsely distributed spatial samples. Descriptive, statistical models are necessary to fill in gaps between such sparse samples data, to estimate locations where sampling should be expanded, and to infer locations where species might be able to live, but currently are not found. Such modeling approaches are increasingly important for resource managers at local-to-regional, and even global, scales.

Complementing descriptive (statistical) models with mechanistic (process) models

Construction of such descriptive, statistical models can also provide some insights into understanding why species live where they do. However, they may lack critical process-level (functional) information, such as CO_2 physiology, to accurately forecast how species might react to or migrate through a landscape or region under, for example, rapid climate change. Thus, the use of these kinds of descriptive models should be complemented by including physiological, phenological, and other life history information in order to understand the processes or mechanisms that allow a species to exist in a given region or locale, in the context of complex interactions with other species and disturbances, and how that species might move from one locale to another.

Distinguishing descriptive and mechanistic models

This life form and physiological information can be used to inform mechanistic models of ecosystem processes and dynamics. Dynamic General Vegetation Models (DGVM) represent one class of such models. DGVMs combine the capability to simulate the distribution of vegetation and its temporal change through successional processes and disturbance regimes. Under conditions of rapid climate change, it is anticipated that different taxa will move across landscapes at very different rates. Rapid migrants often tend to be early successional species having suites of characteristics that adapt them for rapid dispersal, establishment, and early successional capability, following a disturbance. More sedentary species often tend to be late successional and/or locally endemic. Thus, rapid migrants will likely overrun species less adapted to migration, potentially placing narrowly distributed species in direct competition with "invaders" from neighboring domains. This also implies that there will be lags, or disequilibria, between the changes in climate and the geographic distributions of species.

Descriptive models cannot capture such lags. Indeed, a statistical relationship developed between a species distribution at a time when, for example, that species migration lags behind changes in climate, may produce misleading results. *[Counterpoint-reviewer's comment: Purely statistical trend analyses are designed to estimate rates of change, and may do so better than poorly parameterized or unvalidated process models. Spatial/migration lags can be estimated by using proximity as a predictive variable in a regression-like analysis.]* Thus, planning for the eventual shift from descriptive to functionally mechanistic models will be important to accurately depict both current and potential future species distributions. In addition to the dynamics and lags between species and the environment, changes in CO_2 concentration (expected as a consequence of climate change) can alter the relationship between species and environment, rendering statistical models that are accurate under current conditions, inaccurate under future conditions. For example, elevated CO_2 can alter the photosynthetic characteristics of a species and its water use efficiency, and increased water use efficiency may allow a species to live in drier climes than it currently lives in.

Descriptive models generally simulate the "realized" niche of a species, which could be reduced in spatial extent from the "fundamental" niche space by competition, predation, herbivory, or other disturbances. Under a rapidly changing climate all of these interactions may change, such that the future "realized" niche might be different than that under current conditions. Also, the realized niche under dynamic change may be quite different from that under some possible future climate. Species will likely re-organize into different mixtures or communities, affecting the larger "emergent" biodiversity patterns and the resulting ecosystem processes. Mechanistic models have the potential to more accurately simulate these dynamics and possible future states. *[Counterpoint-reviewer's comment: But these are also interaction terms in a conventional descriptive analysis, and depending upon the quality of the data and validity of the reductionist model, could, in principle, work as well or better. It is usually an empirical matter of which kind of model produces a higher r^2, if that is an appropriate way of measuring success.]*

Laying the conceptual foundation for future generations of biocomplexity-capable models

In the same sense that simpler statistical/descriptive models may eventually become complements to, or be replaced by, evolving mechanistic/functional models, those functional models too may subsequently be supplanted by other, still more capable and representative models. Likely those next steps in modeling will respond to challenges to address biological complexity, that is, the inherently complex behavior of living organisms operating as individual agents dynamically interacting in and with ecosystems. The need to address the related limits of mechanistic model projections may become most obvious where multiple simultaneous species interactions and real-time feedback from environmental (co-)modifications must be considered. These challenges will be more provocative where growing human involvements become implicit considerations in efforts to model complex interactions in and with living systems. Because these kinds of biologically complex system relationships are not strictly computable ones (within current paradigms), they will require other modeling approaches. One example is agent-based simulation which relies on behavior or decision rules applied through cellular automata representing individuals interacting dynamically and iteratively (ecologically) to produce a range of plausible, potential outcomes.

To succeed, innovation will also be required in acquiring the additional dimensions of data and information needed to drive and qualify such models of complex behaviors in living systems. Just as purely taxonomic species data must now be augmented to include functional attributes in order to more fully realize mechanistic modeling capabilities, subsequent generations of models will also need to be enhanced. They will require learning how to structure supporting data in ways that capture and preserve revealing information about the ecological context (biotic and abiotic) in which observations and measurements about species and organism behaviors are made. If means can be found to capture adequate contextual or semantic, ecological information about the circumstances surrounding and accompanying core data observations, it will then become possible to recursively or iteratively re-examine those data. As modeling capability and knowledge expands, it will be possible to better understand their meaning and reinterpret their implications. Likely such future modeling capability will rely on projecting multiple possible outcomes in an attempt to bracket a range of potential future scenarios and patterns, rather than predicting singular forecasted results or consequences for complex biological systems. The implications of this challenge for data and information acquisition and handling will not be trivial.

As statistical models are further refined and adapted, and as mechanistic models continue to evolve with the functional data needed to properly implement them, thought will also need to be given to new generations of models that will better anticipate future states and trends involving biologically complex systems. Parallel effort will need to be devoted to the peculiar data and information requirements of these next generations of models.

Recommendations:

1. **Implement a clearinghouse for existing biodiversity models**

There is currently an unmet need to develop a clearinghouse for biodiversity models to support both research and decision-making. Such a clearinghouse will provide a location to house, and capability to distribute, biodiversity models with guidance for their selection.

The NBII should develop this capability in concert with both the biodiversity modeling and the management/decision-making communities to ensure that the needs of both communities are designed into clearinghouse functions. The four actions below propose the functionality that should be incorporated into an effective clearinghouse implementation.

Proposed constituent actions:

a. **Build classifications of critical ecological (functional) attributes of species**

Anticipating future states under changing circumstances

As noted earlier, rapid environmental change is now the rule rather than the exception and consequent impacts on plant and animal biota are ubiquitous. Modeling to simulate these changes in order to understand their resulting impacts on both current distributions and future dynamics of species and ecosystem processes suggests the need for comprehensive modeling at all scales from site to landscape, region, continent, and globe. But the number of species, even at small scales, is sufficiently large to preclude the possibility of modeling everything everywhere. At larger scales, the problem becomes even more difficult, if not intractable. With proper data and classification schemes, modeling can provide a means to use limited knowledge about (existing) target systems to anticipate their future state(s) under changed circumstances.

Exploiting functional similarity to simplify challenging modeling tasks

Fortunately, it is in the character of biology that evolutionary processes have found similar (but non-identical) solutions to the problems of survival in similar climates, but in very different geographical domains. A good example may be seen in the convergent evolutionary pathways and consequent similarity between Cacti in the Americas and Euphorbs in Africa, where each occupies an ecologically similar but geographical disparate domain. They are able to do this by exploiting different evolutionary "strategies," drawing on different functional traits. Other examples abound and are borne out by studies showing the functional similarities between northern and southern hemisphere biota even though their genetic relatedness is very low. Modelers of vegetation dynamics at large spatial scales can take advantage of these functional similarities to simplify challenging modeling tasks by simulating distributional changes using aggregate Plant Functional Types (PFTs), rather than species.

Functional indicators for invasion, dispersion, and migration

In earlier modeling efforts, PFTs were very highly aggregated, such as "Boreal Evergreen Needle-leaved Tree" or"Temperate Deciduous Broadleaved Tree". These PFTs are then simulated in dynamic interaction with each other, including overstory and understory interactions between, for example, trees and grasses. The modeled ecosystem is then classified into a "physiognomic" community, such as "Temperate Deciduous Forest" or "Tropical Drought-deciduous Savanna." This technique proved valuable in climate-change modeling, for example; but these early and rather simplified depictions of PFTs are no longer adequate for dynamic simulations of rapid climate change. There is now a need to be able to simulate the rates of migration of PFTs and their interactions with extant PFTs in the locales to which they are migrating. That is, circumstances in which exotic plants will be invading communities in neighboring domains (in a climate change scenario, for example). To accomplish this, existing PFT descriptions must now be subdivided into finer descriptions that capture variation in dispersal rates and other associated functional characteristics. It is known, for example, that pines can be arrayed along a gradient from rapid to slow dispersal capability. More fortunately, a host of other traits appear to co-vary with dispersal capability, such that whole "syndromes" can be defined within PFTs, ranking species from more to less invasive. These syndromes are similar in concept to "early successional" versus "late successional" plant classifications.

Expanding from taxonomic to functional classification systems

Most current plant classification systems are based on genetic relatedness, that is, they are taxonomic (only). A new complementary system of classification or a revision of the current system based on functional attributes is needed for the further development of dynamic vegetation models. Such a classification system is beginning to emerge, but it is developing slowly and in an ad hoc manner. As noted earlier, newer DGVMs are based on measured physiological and other life history characteristics of plants. The advantage of this is that these models can now be implemented over small geographical domains with their more detailed PFTs specified (parameterized) as actual species allowing direct comparison of "simulated" and "observed" species distributions. If a species is simulated by itself, then a map of that species' "fundamental" niche (distribution) should be possible to produce. However, when a species is simulated in its competitive milieu (shared context) with a number of other species or PFTs, its "realized" niche can be simulated. Such relationships can be used for improved model construction and validation. For application to larger spatial extents, PFTs can still be defined more generally to examine the dynamics and potential shifts of, or impacts on, ecosystems over continental and larger geographical domains.

Construction of a more detailed, functional classification of PFTs will allow extension of the NBII's existing physiognomic classification scheme to become a more useful functional description of ecosystems. The bottom of the classification system might culminate in small collections of species or ecotypes. Intermediate to higher levels of PFT classes might contain increasing numbers of species, but those species may or may not be genetically related. In this scheme, any given taxon of related species might recur (be classified into) several different PFTs. Although this discussion is focused on plant species and PFTs, it should also be possible to build a similar hierarchical PFT classification system for terrestrial and aquatic animals.

[Reviewer's comment: A new classification scheme to capture functional attributes of vegetation classes may be a very complex exercise and in any case would have to coexist with existing vegetation maps in most modeling

environments. Function-based classifications would require the development of a multidimensional (structure, energetics, water relations, biogeochemistry...) theoretical construct (analogous to species concept and genetic linkages of taxonomic classification) on which to base the classification. While such schemes would undoubtedly be useful, an interim solution may lie in ability of exiting and new classification schemes to contain at a minimum descriptions of vegetation that enumerate functional/ mechanistic as well as taxonomic properties. The ability for the NBII to support functional modeling lies in its ability to capture important attributes of vegetation that help predict what species are present in the vegetation and the functional role of the species.]

The NBII should provide leadership in a collaborative effort to extend the existing physiognomic classification system to include functional attributes, as it did when it helped lead the development of a national system for classifying vegetative communities and associations. Products of these collaborations should be delivered through the NBII's biodiversity modeling clearinghouse.

Other classifications options – Life Histories
Another promising direction in classifications to support biodiversity modeling is in the parameter space of species' life histories. Recent theoretical work by Hubbell and others suggests that some important properties of species-abundance and species-area distributions can be captured by community ecology models characterized as "neutral" with respect to species ecological and demographic differences, but that do depend on the aggregate properties of reproduction, survival, and dispersal. Understanding the distributions of species' life history parameters and the interaction of these parameters in models like Hubbell's may provide a useful path to classifying organisms for assessments of biodiversity.

b. Establish a decision support system for selecting and using biodiversity models

Need for multi-criteria matching of problems to modeling requirements
Two very different and complementary types of models are distinguished here. The first type, statistical models, provides a descriptive analysis of biodiversity (species distribution for our immediate purposes) over the landscape. This type of model summarizes or estimates where species may occur in landscapes based upon purely statistical means. These models do little to address functional questions. The second type, mechanistic models, attempts to directly address ecological function. It attempts to ascertain why species may be present or to understand their role in the landscape or ecosystem. In practice, most working models include aspects of each approach. Within each of these two classes of models, however, there are model differences based upon assumptions, data, and a variety of parameters. To reduce resulting confusion and to better inform prospective users of such models, a decision support (selection) tree may be necessary to aid users in deciding what models may be most appropriate to investigate or address a particular problem. A decision support tree for selecting and using biodiversity models needs to incorporate information in a number of dimensions. By determining the position (requirements) of a proposed study among these various dimensions, a choice of models can be made that is most appropriate to the problem under study.

Model output needed (questions to be answered)
Most important among these dimensions (considerations) is the nature of the question that the model must address. One way to specify such questions is through classification of model outputs. Statistical models provide outputs with organismic content. These include spatial distributions of individual species; distributions of assemblages; distributions of diversity measures such as species richness, evenness, or any of the many other measures of diversity; distributions of population parameters such as individual species abundances; and distributions of species habitat suitability indices. Functional models provide outputs that represent functional properties of organisms such as physiognomic, trophic, or life history classes. Other types of output include energetic or other material properties of ecosystems such as productivity, biomass, or components of system biogeochemistry. A different type of output is that used by geographic targeting models for selecting conservation, restoration, or other sites for management or research purposes.

Spatial framework
Another dimension of model selection is the spatial framework needed for the study. The extent of the study area is the first defining property of the spatial framework. This extent may range from global to continental, regional, or down to a specific management unit, or it may require multiple scales. The spatial resolution of the study is also important in defining model selection. Some models use regular, geometric, cellular divisions of space and others use arbitrary divisions defined by political, ecological, or other physical criteria. Although it is relatively easy to convert (translate) between different spatial frameworks,

some models may not have this capability themselves, and may thus require spatial data preparation (pre-structuring) prior to modeling. Some models may incorporate or require a hierarchical spatial structure. Many of these issues are at least partially addressed in emerging Open GIS technologies, and current model developers should build architectures for spatial data that fit the NBII's strategy for migrating to Open GIS tools and standards.

Temporal framework

The temporal framework is another important model specification. Many models output a single point-in-time estimation for change over time. This type of model would be used to do estimates for the boundaries of the time extent under study, and any intermediate points in time. The time extent could include any combination of past, present, and future points. Time dynamic models could include continuous time domain models such as ordinary differential equations (ODE) models, or discrete simulation models with either regular time steps or event-driven time steps. Some models may incorporate or require a hierarchical temporal structure.

Input data / information

Another dimension of model selection is input data/ information requirements. A wide variety of parameters are needed for models ranging from spatially or temporally explicit data to system-wide properties of organisms or ecosystems. Some models require initial distributions of, and/or other information about organisms from which they then build estimates of more comprehensive distributions (see "Expanding from taxonomic to functional classification systems" above). Many models require non-organismic, environmental data such as climate, topography, soils, hydrology, and others. All such data are now often required in a spatially explicit format in regular or irregular spatial frameworks. Some models may provide capabilities to convert input data to model formats, but many may not. A related issue is whether model inputs need to, or should, conform to prescribed data standards, or conventions or common practices in the relevant content community, and whether such conformity is, or should be, documented in model metadata. In practice, it is unlikely that data contributors can or will convert data into common schemas or formats, because many data sets are gathered for purposes different from biodiversity applications, or are constrained by long-term investments in incompatible legacy methods and protocols. An alternative and perhaps more realistic goal is to promote wide use of "Semantic Web" technologies under which schemas or ontologies for heterogeneous data are declared and expressed in open-source formats (such as Extensible Markup Language [XML], Resource Description Framework [RDF], or Web Ontology Language [OWL]) in ways that unambiguously define the meaning of the data and permit machine access and interpretation across multiple applications and platforms.

Computing resources

Although appropriate computing resources continue to become more widely accessible in the marketplace, variation in computing capabilities will certainly persist among potential modeling users. Modeling requirements for storage and compute cycles will be important for some users with limited resources. One dimension of a decision support system needs to balance these resources and related constraints. Included in this dimension could be the possibility of a hybrid solution through distributed implementation of models. This will, in turn, require consideration of whether the model can be downloaded to a user's computer, the type of programming languages used, and other system requirements for models that can be so distributed. A decision support system could also include results of the performance of models solving similar problems (not to be confused with data benchmarks referred to in "Accreditation" below).

Accreditation

Another aspect of model selection is the accreditation status (confidence in and acceptance) of models. By this we mean any certifications or indications of the proved or approved status of the model within the content community that it serves. Accreditation could take such forms as peer-review reports by the NBII or other panels (see "Enable and promote peer review of biodiversity models" below), accuracy performance measures (benchmarks) on standardized data sets, or anecdotal evidence documenting author or user experience with the model (satisfaction or dissatisfaction). Researchers in the computer science community are experimenting with automated methods for measuring "trust" in Web (or other) services. Google rank is a simple example. It is likely that the NBII can adopt methods using the nature of calling links and services to automatically generate ways to help users determine the reliability of, and appropriate uses for, competing models.

Uncertainty assessment

Because no model can be perfect, assessments of model output uncertainty may be an important feature for some users. For complex space-time dynamic models with stochastic components, such assessments may be impossible in any comprehensive way. However, Monte

Carlo simulation of the effects of variation or error in input parameters may be possible. This kind of uncertainty or model sensitivity analysis could be applied to less-dynamic models as well. A dimension of model selection, then, could be capabilities or amenabilities of models for such uncertainty assessment. Note that expressing and visualizing uncertainty in a complex spatial setting continues to be a research challenge.

Method and approach

Of course, an important component of model selection is the modeling method to be used. Part of the goal of a decision support tree is to suggest methods that are appropriate for studies with given output, spatial, temporal, computing, or accreditation requirements. This sifting through the dimensions of a modeling problem should help narrow and defend the choices of potential model users who have little familiarity with the universe of models that may be available. Even so, there may be a number of methods that can meet their requirements. A useful first step in distinguishing models is to discriminate between those that use assumptions about critical model parameters in their structure (parametric) and those that do not. Models that could be included in the parametric approach are those based on geostatistics, or other regression techniques such as multiple regression or decision trees. Non-parametric approaches include those using such modeling methods as genetic algorithms, neural nets, or simulated annealing. One example would be the Genetic Algorithm for Rule Production (GARP) modeling environment. Another approach is to choose models that perform space-time dynamic simulation of ecological processes or mechanisms in order to estimate their outputs. These often complex models may include some distributional assumptions about their inputs, but the relation of the outputs to the inputs is usually highly non-linear and possibly chaotic. For some kinds of modeling requirements, space-time dynamic simulation models may be the only possible method.

Advancing from less informed to more informed model selection and use

The NBII should organize a collaborative effort to develop a decision support system to guide the selection and use of biodiversity models. This system can be used as an (educational) organizing element for models contained in the clearinghouse, will facilitate access to necessary metadata to characterize models, and will enable more effective and reliable search and retrieval of relevant models by users. Decision support tools should help modeling users identify what class of models are best suited to address their problem and to facilitate screening for those specific models whose requirements (dimensions above) can be best met. *[Reviewer's comment: The best fit of competing models to particular applications must consider simultaneous constraints of data availability, scale, time-scale, computer power, need to address policy or legal constraints, and so forth. It is really smarter metadata. These might start by developing ontologies for models, other services, and data sets. Software "agents" could then automatically explore available data and services for fit to a particular query of class of questions.]*

c. Track reuse of biodiversity models

Even a simplistic clearinghouse implementation might, with modest effort, track by whom and when a particular model format or version was accessed or downloaded, but closer tracking would likely be advantageous. One would want such a clearinghouse to monitor and communicate who is applying an NBII-hosted model, for what purpose they are using it (objectives or questions to be answered), and the model's status relative to its intended uses and the author's maintenance of it. Note that if such applications are on the Web, emerging "Semantic Web" tools may both do some of this automatically, and use the results to create and update information on appropriate uses and trust, to better guide new users.

Fitness for use

There is a need for the exchange of more systematic information about model specifications and requirements, about the life cycle of models offered/available, and better information about users and their purposes and experiences. Adequate tracking information will help to create an understanding of the variety of model uses that the NBII is supporting, and would address concerns about whether a model was being applied for uses other than those intended by its author. Models can sometimes legitimately serve applications for which they were not originally intended, or they can be fairly adapted to do so. While some of these uses may be considered "extensions" of their value, others might be considered misuse or inappropriate use of a purposive modeling approach. Equally important, tracking data would also help to determine what models are most widely used and why. This information could then be used to improve NBII services, and could be fed back to the modeling community to guide the improvement of models and to inform modeling research planning.

Revision and expiration

Tracking data can also help by properly accounting for the life cycles of models. Biodiversity models may be relevant and appropriate for use within a finite term, but they can quickly become out-of-date if they are not maintained or updated as new applications and approaches are developed. When submitting a model to the NBII Clearinghouse, a provider may stipulate, for example, that a model was posted on a given date, that the expected life span of the model expires after a specific period of time, and what may or may not be expected for the model's further development. After a model's life cycle expires, the model may be removed from the service (access at the clearinghouse), or the provider may offer a more current replacement model or extend the expected life of the current model by revision and reposting.

Updating models through revision must be facilitated by versioning, that is, by providing a version identity code or number for each rendition of a developing model. By enabling (or allowing) multiple versions of a single model, it is expected that researchers may be induced to make available works-in-progress. This flexible approach to version tracking is important because biodiversity modeling is a rapidly evolving field and many models remain in draft form without ever reaching a point that would be identified by the author as a final version. Note that to the extent that models are used to develop data that are subsequently used in downstream analyses or applications, it is important that the model version used be archived, so that dependent results can be documented and replicated.

The NBII should track specifications, development, and use of biodiversity models contributed to its clearinghouse. As proposed here, tracking will foster the creation of new and improved biodiversity models that effectively address the key issues for which clients seek models. Also, by tracking requests for models, as well as provision of models, the NBII can better understand unmet and evolving needs of biodiversity modeling applications and their users.

d. Enable and promote peer review of biodiversity models

Models, like publications, are representations of new knowledge (experience) gained from application of basic scientific principles. And like publications, models can have a large impact on scientific advancement and they may serve as tools to inform biodiversity management decisions. Accordingly, models require the same care and treatment as other scientific tools and products. In current practice, models receive a peer review only when submitted as part of a broader paper for publication in a scientific journal. There are no separate review processes that deal specifically with models, or more particularly, with biodiversity models.

The NBII should proactively work with the biodiversity modeling community to develop a process in which the work of model authors is reviewed by one's peers. Such reviews should be accessible at the clearinghouse and perhaps in an electronic journal, also implemented by the NBII (see "Provide education on information management and modeling" below).

This review process could be the first step in posting models to the NBII biodiversity modeling clearinghouse and would be an effective way to encourage submission of models for use by others. Delivery of a peer review process would also be an excellent product for the new World Data Center for Biodiversity and Terrestrial Ecosystems, co-located with the USGS-CBI. Useful subjects of such peer reviews might include metadata, input data, model output, descriptions of independent variables, modeling methods and versions, model validation techniques (if any), data on model accuracy (especially errors of commission and omission), and data and version stamps.

B. Data and information for biodiversity modeling - status and expectations

Data are required, variously, to invent, adjust, verify, refine and evaluate models. Thus observations and measurements (data) are necessarily linked closely with "predictions" and estimations (models). Biodiversity modelers begin with current scientific understanding of a particular species distribution pattern and biological processes, and couple those with known elements in the environment to test basic understanding of phenomena or to "forecast" future species distributions. Normally future-ward modeling may also be done in (or may require) a retrospective operation called a "hindcast." Thus reliable modeling requires ready access to good quality data and information that accurately represents what is already known about problems, qualified by what was expected or learned from prior modeling efforts. The importance of adequate, credible, accessible, well-documented data and information cannot be over-emphasized if robust results are required from biodiversity models.

Recommendations:

2. Identify and catalog the relevant data universe

The NBII should directly address the immediate data and information challenge to better serve biodiversity modeling by making biological and environmental data accessible to modeling tools, and by making the results of modeling processes available to a variety of users, including scientists, decision-makers and the public.

The first step in achieving this goal is to identify the scope and nature of the pertinent data universe, particularly in regard to biological data resources. General categories of data needed for modeling purposes include organism occurrence (such as specimen and observation data), organism and population abundance (such as survey and population monitoring data), and environmental characteristics (such as climate, land-use, elevation, vegetation/habitat). Although some models require a broad array of data elements, in many cases only a few parameters are needed. Minimally, biological data must reference a taxonomic context (via taxon names and concepts), and then link to environmental data via a spatial (geographic) coordinate system. Model output itself should also be considered to be data, and may represent input data for use in other and/or subsequent models.

Constituent actions:

a. Automate competent data discovery

Enumeration and description of data
Simple enumeration and description of data resources enables potential users to discover those resources and evaluate their fitness for a particular use. The NBII's Biological Extension of the Federal Geographic Data Committee (FGDC) Metadata Standard and its Clearinghouse Mechanism have made significant progress toward this end. The NBII metadata catalog should now be analyzed to identify gaps in taxonomic and geographic coverage, and in types of data available. While it is certainly a necessary step to provide intelligence about existing data sources, putting those data to use may be hampered by the often laborious task of transforming them into a more useful structure.

Beyond simple human-readable metadata
To allow more efficient access to distributed data, electronic documentation of data must go beyond simple human-readable, descriptive metadata. Ideally, metadata for each data source should be documented with sufficient detail and consistency that data element cross-walking can be achieved more or less automatically. Without reliable automated data element cross-walking to relate disparate data sources, data users (such as members of the modeling community) are burdened with the manual (and possibly duplicative) implementation of this task. The panel encourages the NBII's involvement with existing efforts to achieve this level of automation, such as that undertaken by the Science Environment for Ecological Knowledge (SEEK <http://seek.ecoinformatics.org/>) and the Semantic Prototypes in Research Ecoinfomatics (SPIRE <http://spire.umbc.edu/>) projects. SEEK tools include Ecological Metadata Language (EML <http://knb.ecoinformatics.org/software/eml/>) and applications to the Long Term Ecological Research (LTER) network (e.g., see <http://www2.bren.ucsb.edu/~rbose/pubs/200005_LTER/Metadata_ppt.pdf>.

b. Assess information needs and data gaps

Appropriate and credible data
Common wisdom holds that models are only as good as the data used to generate them. Biodiversity modeling requires biological, environmental, and ecological input data that are reliable and specifically suited to meet related research or management objectives. Properly satisfying the need for such data means that one must assess the requirements of the modeling community to identify critical data gaps. Filling critical biological and environmental data gaps will prove invaluable to the development of better biodiversity models.

Profiling biological data gaps
Biological data are most commonly available in the form of species collection and observation records. Although their availability is increasing, these kinds of data are still lacking for some taxa and some geographic regions. A concerted effort is needed to evaluate the availability of collection records and other biological data sets nationally and internationally. Results should identify biological data gaps, and at a minimum, characterize them by major taxonomic group, spatial grain of data capture or storage (such as area covered by collection event or method of georeferencing), and data type (presence vs. presence/absence vs. abundance). Priorities should then be established for the order in which these gaps should be completed (filled) through expedited cooperative activity.

Profiling ecological and environmental data gaps
The relative distribution of plants and animals (one indicator of the diversity of biota) is determined by a wide array of environmental factors operating along a continuum of spatio-temporal (space-time) scales. Over the past two decades, there has been a great increase in the development and availability of geographically or spatially-referenced abiotic environmental data (such as climate and hydrologic parameters, elevation, and topography). This

trend complements the increase in collection and availability of biological data. Notably, however, the development of environmental data has been influenced by a variety of user needs, and not necessarily the needs of the biodiversity modeling community. Surveying available geospatial data and other data required by modeling practitioners could help identify critical gaps to be addressed by future development of environmental data for accurate bio-physical models. Such a survey should account for the availability of required environmental data relative to major feature categories, their geographic extent, and spatial grain.

Synthesizing data for ecosystem management
Measuring and mapping the spatial distribution of species and community types across landscapes is just the first step toward conserving biodiversity. Understanding these patterns, the underlying processes that shape them and potential human-induced alterations of them must be our ultimate goal because such understanding is essential for effective ecosystem management. Understanding underlying ecological processes requires knowledge of the ecology or life history of the species we are attempting to conserve. There have been many independent and widely scattered efforts to characterize species according to critical ecological attributes (such as reproductive guilds, physiological tolerances, plant life forms, and so forth) that determine their distribution and abundance across the landscape.

Closing data gaps
There exists no concerted effort to generate standardized characterizations for a wide range of taxonomic groups, and important structural and functional features of the biota of our nation or the world. To close this gap a major collaborative effort is needed among state and federal agencies and research institutions to synthesize widely-scattered and incomplete knowledge into a more centralized, hierarchical schema for classifying species, one that accounts for their key ecological attributes. Formulating independent or collaborative funding and research initiatives to support local, regional, and national biological survey or environmental data development efforts would necessarily lead to enhancements in biodiversity modeling efforts. In addition, a synthesis of our existing knowledge of critical structural and functional attributes of our nation's plants and animals would provide the key to moving from "simple" mapping conceptions of biodiversity to more formal understanding of the processes that drive the creation and destruction of this diversity.

The NBII should facilitate the filling of priority data gaps through several initiatives including: conducting surveys of data needs and data availability, hosting a forum for exchange of data and information through regional or national conferences or via the Web, facilitating the creation/enhancement of environmental and ecological data sets by serving as a coordinating entity and a potential funding source, and operating as a broker to discover other/new funding sources and to mate them with data projects requiring additional resources.

c. Advocate and fund data digitization
Digitizing non-digital data, and transforming existing digital data into more usable formats (such as geo-referencing textual locality descriptions into spatial coordinate systems) still represent the most fundamental challenges to using many existing data resources. These data-availability problems exist in both the biotic and abiotic (environmental) data domains. The NBII should work actively with its partners to identify non-digitized data resources, to acquire needed funding, and to facilitate digitizing these data.

3. Improve data integration through schema element or ontology reuse
Improved access to biodiversity data and information relies upon greater standardization of data structures (or at least standardized external views) and formats. It also requires standard mechanisms to describe data and standard programmatic methods for accessing those data. The ultimate goal of this standardization is to provide more nearly seamless integration to users across sources. The NBII's biodiversity modeling clearinghouse should adopt the goal of standardizing of schemas and/or ontologies, or more correctly, promoting ontology and schema element reuse, to help improve the current state of disparity (limited compatibility) between existing data sets. Similarly, there is a need to develop a multiplicity of data and model service interfaces, the programmatic interfaces used by machines to connect with data and simulation models, or other programmatic services.

<u>Proposed constituent actions:</u>

a. Develop and manage a federated data schema

End members and current practice in data integration
The simplest approach to begin schema standardization is to build a catalog of metadata relevant to the available data sources (such as the Biological Extension to the FDGC Metadata Standard, and the NBII Web Resources Catalog), without committing effort to cross-walk specific data elements. As stated above, this approach is necessary for

documenting the data universe, but it does not address the labor-intensive step of combining data resources on a case-by-case basis.

At the opposite extreme, the alternative is to develop a single, all-encompassing standard data schema, to which virtually all existing data sources can (and must) be mapped. This approach has had some success when applied within a specific subset of data sources where control and/or willing compliance is high. For example, the museum collection community is collaborating to develop networks of distributed data resources that are based on a federation schema. A federation schema can be thought of as a set of generalized, but still well defined, data elements common to a broad range of otherwise different data resources. By having data providers map (conform) their heterogeneous resources to a federation schema, a large amount of work is distributed across a broad population of data managers.

Systems such as the Species Analyst <http://speciesanalyst.net/index.html> use federation schemas to integrate conceptually overlapping portions of data resources. The Darwin Core <http://tsadev.speciesanalyst.net/DarwinCore/darwin_core.asp>, which underlies the Species Analyst (and now Distributed Generic Information Retrieval [DiGIR], see below), is a simple schema. It applies to a comparatively large number of records because it is relatively abstract, but it is insufficient to represent (and transport) data concepts that apply to more specialized data sets. However, when considering the much broader domain of biodiversity and environmental data resources of potential interest to the NBII and the modeling communities it can serve, history and experience suggest strongly that the development of such a monolithic data schema to accommodate all (or almost all) data sources unlikely to be successful, at least within the foreseeable future.

The alternative – an adaptive, practical approach to data integration

A better solution lies between the extremes of building either a simple catalog or one all-encompassing standard data schema. Rather than devising a single monolithic schema into which all data resources must be mapped (conformed), several (perhaps fewer than one or two dozen) "standard" schemas could be developed that would collectively accommodate the majority of well-known kinds of data and data uses (such as models). As described above, the Darwin Core is one such schema that already exists. Although intended primarily for Natural History Collections' data resources, this schema can accommodate a wide variety of data types, wherein each record represents an occurrence of a particular taxon at a particular place and time. Thus, whereas Natural History Collections' data will map "robustly" (many matching elements) to the Darwin Core, other data types might map somewhat less robustly, but still map well enough to provide meaningful data on taxon distribution over time. Still other types, such as those containing exclusively abiotic geospatial, environmental data, might not map to the Darwin Core at all. Note that a similar approach, ontology languages, under development as part of the Semantic Web, are capable of expressing a greater range of relationships among data elements. It is likely that as federated XML schema approaches migrate to being expressed in next generation Semantic Web forms, some of the difficult problems of "mapping" between differing vocabularies and concepts will become more amenable to automated cross-walking and semantic processing.

A federated solution - multiple standard data schemas (MSDS)

There would likely be substantial overlap of specific data elements among different schemas. For example, fields associated with taxonomic identification, and/or fields associated with spatial coordinates, will likely appear in many if not most of the different schemas. Other data elements specific to a certain class of data sources (such as natural history collection data, population monitoring data, environmental parameter data) would likely appear in only one or several of the standard schemas. As indicated in Figure 1. each data source could potentially conform (map), to a greater or lesser degree, to more than one data schema. (Note that bold lines in the diagram represent the most robust mappings; dashed lines represent the least robust mappings.) With the right set of standard data schemas, each of which is intended to address a particular class of data resource, at least one, if not multiple standard schemas, could accommodate most data.

Implementing MSDS

The rationale for establishing these multiple standard schemas is to reduce the plethora of protocols among existing data sources and their respective schemas into a more manageable number of alternate schemas. As opportunity and need allows, individual data sources should be manually mapped to one or more of the standard data schemas. Once a data source is mapped (conformed) to a standard schema, it will subsequently become available via that standard schema for all future users. When a user needs access to data, instead of simply identifying potentially hundreds or thousands of disparate data sources

Data Federation

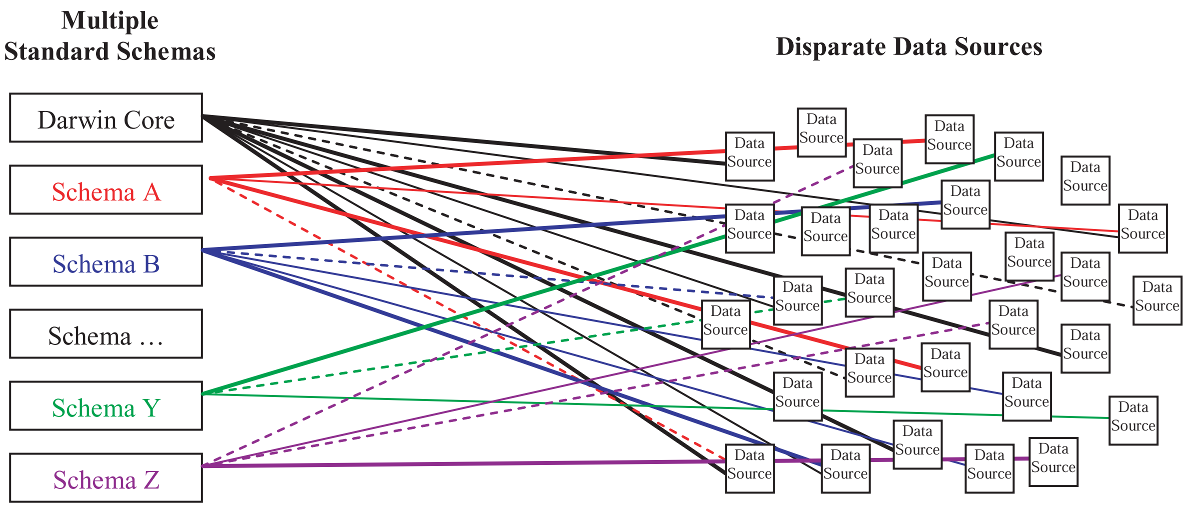


Figure 1. Relation between multiple data sources and multiple standard data schemas

and manually conforming each one to a single schema, the user need only identify which one, or which of several, of the standard data schemas contains elements useful for the intended purpose. Thereby the user gains direct access to all relevant data sets without the need for manual transformation on a source-by-source basis. In this way, schema standardization offers an intermediate solution that, while requiring some human mediation (in the form of mapping existing data sets to one or several standard schemas), also enhances automation of data access (once mapped to a standard schema).

Identifying and developing automated users and support tools

The NBII should review the extant metadata catalog to determine how to cluster data resources into optimal data type, in order to clarify, specify, and prioritize schema development. This should be done in concert with a similar analysis and classification of the inputs required by modeling tools, so that classes of "automated users" or data consumers are identified and used as guidance. (For examples see "Lifemapper" system, <http://www.lifemapper.org/>), based on the Darwin Core federation schema and the GARP modeling tool for predicting species' distributions.) The NBII should further facilitate the development of those various standardized schema, as particular needs or classes of data resources are identified through workshops and other forums for discussion among interested and knowledgeable parties. The NBII can help the process by supporting the development of software tools ("Wizards") to assist data holders in the process of mapping (conforming) their data source schemas to one or more of the standardized schemas. Similar "Wizards" would assist researchers in designing schemas for new data sets in such a way that they maximally conform to one or more of the existing standard schemas and/or identify needs for new schemas or extensions of existing standard schemas. Finally, the NBII should maintain a network for communication with data holders about modifications to existing standard data schemas and the introduction of new standard data schemas.

b. Registry support and data services

Beyond passive discovery and secondary characterization of data and services

A registry system allows an active means to identify the existence and location of both data resources and data services. Active registration of data sets alleviates the need for passive discovery and secondary characterization (metadata generation) of information resources. Also, registration of data services (standardized programmatic interfaces to data resources) allows the reuse of general data services that may have broad application for many data resources and data access needs. In either case, the NBII should adopt a centralized registry based on a common, well-known standard, such as Universal Description, Discovery, and Integration (UDDI), to remove potentially significant development activity required to build a custom registry.

Registration process

Registration of data sets should be a straightforward process that simply involves selecting a schema and access Applications Program Interface (API) from the registry. To support this there should be some sort of validation process that is invoked during the registration process to ensure that: 1) data are accessible; 2) data conform to the selected schema(s); and 3) data values are in the expected range (where appropriate). The NBII should explore options for data resource registration, and seek to either support and coordinate with existing registry efforts or to develop new registry efforts where none exist.

Registration authority and mediation

The NBII's proposed registry of data services should support and document the location of services, description of services, and responsible parties for both services and service descriptions. It is critical that duplication of service descriptions does not occur (such as two service descriptions that describe the same service type). Hence, some authority should be designated to ensure that this does not occur. Registration of new service types in the registry should not be a trivial process and should require human mediation, rather than being an automatic or autonomous process.

4. Promote data quality, data exchange, and peer review

Generation of data schemas for existing data sets contributed to the NBII's biodiversity modeling clearinghouse should be made as simple as possible and should provide significant incentives for participation by data producers. Examples of incentives might include data cleaning or visualization services. Comparison with existing data sets may be useful in some cases and software tools need to be developed to assist this process. These tools are more likely to be adopted if they are designed for cross-platform use, and maintained and improved through an open source process. Tools to assist with the creation of new data sets should be provided to help reduce schema diversity or better, simply promote reuse of existing schema components where possible. Again, cross platform and open source tool development will help encourage broad adoption.

A peer review process should be implemented by the NBII for some types of data sets, such as "standard" environmental layers and critical distribution. A large part of data quality evaluation tasks should be achieved with the kind of automated tools described above. Peer review will help ensure data sets and their metadata are meaningful, complete, and accurate. Peer review of data sets should not be allowed to become an impediment or precondition limiting accessibility of data sets. Instead, data sets should simply be tagged to indicate their peer review status and then offered freely.

5. Provide education on biodiversity information management and modeling

Educational efforts toward implementation of models and modeling should relate to the two general NBII objectives of advancing research appropriate for resources management and providing the results of this research in a format useful for resources managers. Educating end users about the importance of models, how they are developed, their uses and limitations, and their interpretation should be an inherent part of the design of any NBII modeling services interface.

<u>Proposed constituent actions:</u>

a. Use metadata to educate model users

Providing incentives and enforcing requirements for model developers to follow appropriate metadata and documentation standards (specifying the major approaches and assumptions in the models) would not only benefit end users, but would also encourage modelers to consider appropriate ways to explain models to novice or new users. This scheme therefore provides benefit to both researchers and end users by providing model metadata listing key assumptions and approaches pertaining to each model (including model outputs). The NBII can assist managers in assessing the relative utility of a given model for their particular application. Doing so would also be prerequisite to full implementation of a decision support tree that assists users to decide which of many modeling approaches and models might be most appropriate to address a specific question/problem (see "Establish decision support system for selecting and using biodiversity models" above).

b. Develop multi-scale consciousness and ecological perspective

[Reviewer's comment: One challenge for incorporating modeling facilitation services within the NBII arises due to the hierarchy of management.] Managers using models at broad spatial extent may have a quite different perspective about use of models and data than managers who deal with a particular, more spatially-limited location. Local resource managers may not understand why information generated from modeling at broad spatio-temporal extent may be necessary to address what appears to them to be a strictly local problem. *[Reviewer's comment: This suggests the need for different means to deliver education to different*

potential users, although having a national perspective represented even at regional gatherings may be helpful so that regional meetings don't become focused solely on local ecological processes or management issues.]

c. Educate collaborators about biological and ecological informatics

Delivering education associated with the modeling components of the NBII can precede in a number of ways. One may be to collaborate, for instance, with the USGS Research Units or Cooperative Ecosystem Study Units at universities and state agencies to develop biological and ecological informatics courses or programs, or to integrate coursework from existing programs at a university into other programs. Existing programs suitable for this delivery might include those teaching Conservation Biology, Ecology, Biology, Fisheries and Wildlife, and Natural Resources Management, or existing extension programs adapted for state and local managers. This might also include the development of short courses for state agencies that would be run out of, and in collaboration with, the Co-op Units, but involve NBII staff or collaborators.

d. Train and expose users to biological and ecological informatics

Alternatively, the NBII might consider helping to establish training workshops/short courses on biological and ecological informatics at key centers where current research efforts are focused, including universities with expansive research groups in the area of biodiversity modeling, or at nodes with this expertise available in-house. "Road shows" by key model developers who have excellent "success stories" about how modeling associated with NBII data management efforts has aided managers would be useful. These could be held at state agencies and regional gatherings associated with large national societies (such as land-use planners), or co-hosted with organizations such as the Institute for Ecosystems Studies <http://www.ecostudies.org> at the Cary Arboretum in Milford, NY, or held in association with the Council of State Governments, the American Association for the Advancement of Science (AAAS), or the Organization of Fish and Wildlife Information Managers (OFWIM).

e. Develop a teaching text

The NBII should work to help overcome an important constraint on educational objectives: the lack of good texts on bioinformatics that clearly relate modeling and data, particularly in a spatio-temporal context. One remedy would be to encourage authoritative people to compose such texts. Another would be to organize a conference, perhaps associated with such organizations as the AAAS or the Ecological Society of America (ESA), the presenters at which would write chapters to be included in a text designed to advance the field. Again, success stories would be helpful to ensure that the text would meet the broad needs of educating resource professionals about the range of information management and modeling issues. As another potential means to create a biodiversity modeling text, hold a strategy session at the annual NBII meeting on educating those external to the NBII about modeling capabilities facilitated by the NBII.

f. Found and sponsor an electronic journal

The NBII should consider early the influential option of helping to establish an electronic journal (associated with the NBII) that covers issues of information management, biodiversity modeling, and information transfer. Such an e-journal could provide an incentive for individual researchers to participate in an NBII-based collaboratory. If editors of this journal required peer review, it would strengthen the quality assurance in models contributed to the NBII, and provide an additional incentive to researchers who participate. These inducements are valuable because modelers often observe that there are few natural outlets for publishing detailed models. This approach could also build a valuable "research utilization bridge" by helping to overcome both the reluctance of researchers to ensure that their results become formally integrated into (applied) natural resource management, and the reluctance of managers to keep up-to-date with current science.

6. Provide a service outlet to develop models and maps operating through distributed biodiversity modeling systems

In early stages of its action plan (following), a primary concern of the NBII should be to promote and deploy current-practice, interoperable, modeling schemes. These schemes will support the development of systems that entail the automated acquisition of distributed data and execution of models on multiple remote computing systems. This might build upon the infrastructure already provided by the NBII nodes for geographic information systems, adding other functionalities.

<u>Proposed constituent actions:</u>

a. Provide scaled, remote computation service

The NBII should provide a range of computational capabilities at a level of service that is appropriate, supportable, and utilized. Recommended service actions, in order of increasing levels of effort, are:

- Create a File Transfer Protocol (FTP) site for download (distribution) of modeling software.
- Demonstrate implementation of algorithms.
- Implement a test bracket that allows users to check the validity of their (own) implementation of a modeling algorithm based on test data.
- Allow users to run their own data on NBII–hosted algorithms via single user interface.
- Provide a standard middle-ware Web service that allows users to build their own interfaces to the computational service.
- Provide a distributed computational service such as Lifemapper, or other grid-based system, for computational-intensive algorithms.

b. Provide mapping and modeling services via the Web
The NBII should provide a service for end users to develop ecological niche models and distribution maps. This would entail a substantial expansion of the current role of the NBII as a server of data. It would require customization of data and models to suit the needs of users, working at a range of scales from local (such as conservation of rare species) to global/international (such as invasive species threats).

To scale such a commitment, a prioritized list of candidate tasks should be developed for modeling activities to be supported and related useful algorithms. Possible candidates, in no implied order, might include:

- Ecological niche modeling including logistic regression, General Algebraic Modeling Systems (GAMs), GARP, decision trees, neural nets, BIOCLIM, and perhaps co-kriging or other explicitly spatial statistical services.
- Population Viability Analysis.
- Reserve optimization algorithms – incorporating such principles as complimentarity.
- Spatial data service – distributing custom sets of georeferenced spatial data suitable for biodiversity modeling.

c. Build user interfaces using interoperable middleware
The NBII should design clearinghouse interfaces that ensure adoption of modeling capabilities by target user groups. This will require focusing on an interoperable middleware layer upon which a range of interfaces could be built. The Open GIS Consortium Web Mapping Service (OGC: WMS) is an example of this approach in GIS, but could also be addressed in other complex algorithms such as Physiognomic modeling, Population Viability Analysis, and so on. Target audiences and their needs might include:

- Simple stateless interfaces that produce a single image or result,
- Detailed interfaces that expose all parameters, and
- Application-specific interfaces such as for invasive species or species conservation.

The development of such interoperable systems will entail a number of important considerations and will require additional study. Several notable initial considerations are:

- Interoperable data schema or ontologies,
- Development of Web services,
- Standardization of application program interfaces (APIs), and
- Web service registration systems such as the UDDI.

d. Implement distributed biodiversity modeling systems
The NBII could adopt a very central role in influencing the form and evolution of an integrated computing infrastructure to support biodiversity modeling. The NBII has ready access to modeling user groups who would benefit from greater access to advanced computing services, and in return, could review and test the developing components. The NBII also has previous experience in standards-based data development, together with significant expertise in facilitating Internet-based collaborative work that could be used to develop a community dedicated to implementing distributed biodiversity modeling systems.

To accomplish this, the NBII needs to improve its ability to leverage funds in the information sciences though participation in such programs as the National Science Foundation's (NSF's) Information Technology Research (ITR). This might be achieved though outreach to the scientific community by hosting workshops, creating an electronic journal, and proactive involvement in existing, funded computational technology projects and ITRs, such as SEEK <http://seek.ecoinformatics.org/>, SPIRE < http://spire.umbc.edu>, and others.

III. Implementation

A. Adopt terms of engagement and proposed NBII Action Plan - Status and expectations

Workshop participants recognized that successful implementation of recommendations for "A. Modeling the distribution of species" and "B. Data and information supporting biodiversity modeling" would require meeting key preconditions with regard to institutional and management attitudes, commitment, and readiness. Most important in this regard is one underlying criterion for success: that action on the recommendations in this report be approached as not usual or routine, and therefore, incapable of providing lasting benefit unless particular, deliberate, and prior adaptations are made.

<u>Recommendations:</u>

7. Adopt a leadership role in improving biodiversity modeling

It is proposed here that the NBII adopt a leadership role to support and improve biodiversity modeling. This role of facilitating biodiversity modeling through collaborative development and application is a new role not previously filled by the NBII or any other agency or authority. Thus it is one that will require effective interaction between the NBII and a host of other individuals and institutions operating in one or more teams designated according to the tasks outlined in the following action plan. Anticipating this new set of operational requirements and planning to properly accommodate them should be accomplished by a structured, proactive, and flexible response by the NBII.

8. Proactively and systematically engage the biodiversity community

Successful support of biodiversity modeling enterprises by the NBII will require careful assessment of modeling requirements established by relevant communities of modelers and user, and by participating communities of scientists and decision-makers. These requirements will need to be periodically reassessed relative to the strategic and institutional opportunities and constraints of the NBII and its partner agencies. The optimum match of modeling requirements to institutional opportunities/constraints will be most successfully translated to effective action if two conditions are met: making an explicit and overt effort to determine how the NBII may establish good practices that are specifically suited to the support of modeling enterprises, and recognizing that such activities are new ones that may not be fully congruent with, or optimized to, existing institutional practices and procedures at the NBII.

9. Adapt NBII institutional interfaces to ensure successful leadership

Following are specific "criteria for success" recommended as necessary preliminary steps toward successful NBII implementation.

a. Designate leaders, managers, and ombudsmen

Advocates should be named who will be imbued with both the responsibility and the authority to ensure that all reasonable efforts are made for successful implementation of the recommendations and related actions described here.

b. Make NBII's contact persons apparent and accessible

Knowledgeable persons should be known, or be made easily identifiable, both inside and outside of the NBII and its partner organizations. All practical effort should be made to avoid repeated rediscovery of persons appropriate to handle inquiries and problems.

c. Define responsibilities and chain of authority

Routing of requests and inquiries, as well as pathways toward solutions and decisions, should be predictable and widely known as standard operating procedures within the NBII and its partner organizations. Awareness about, and revision of, these procedures should be continual and routine.

d. Orient staff and management with ongoing internal education efforts

A high level of understanding about enterprises implemented to support biodiversity modeling should be achieved and maintained, recognizing that such commitments will be new and constantly evolving.

e. Establish, monitor, and implement strategic and tactical plans

Planning should be proactive, persistent, and adaptive. Business as usual will not provide adequate management support.

10. Begin three-year start-up phase with proposed schedule of effort

Workshop participants concluded that the recommendations in this report would be more valuable, and perhaps easier to implement, if some sense of priority and relative effort were attached. Participants did not conclude that ranking recommendations by order of importance would be feasible or desirable at this early stage. Priorities are expressed in terms of the year in which each recommendation should be started and its period of duration. No calendar dates are suggested, since progress will need to be measured from an as-yet unknown inception date to be determined by the NBII. Instead, start dates are shown by"project year." Similarly, levels of effort that might be expected to be invested to achieve satisfying results are shown as day, month, or year units of effort by individuals, teams, or groups. In some cases it is implied that authority be invested in one person responsible to supervise and accomplish a designated task (recommendation). It is assumed that this schedule of effort will be revised through time to reconcile resources and constraints. However, it is strongly suggested that revision be accomplished comprehensively and systematically, with consultation by a core group of workshop participants, rather than incrementally and autonomously by institutional managers. There is danger of defeat by attrition if any change does not anticipate the scope and consequences of all changes.

<u>Proposed schedule of effort:</u>

Revise/refine biodiversity modeling problem statement

- Year 1: 1 day team effort
  Generate goals and objectives of the NBII pertaining to modeling
- Year 1: 1-3 months team effort. Review report from modeling workshop and formulate a strategy that outlines subsequent goals and objectives.

Define universe of biodiversity models and modeling activity

- Year 1: 3-6 months. To be done in conjunction with identifying goals and objectives.

Assess information needs and data gaps pertaining to biodiversity modeling

- Year 1: 3-12 months, individual or team effort. Conduct surveys of data availability and data needs.
- Year 1 and/or 2: team effort. Provide forum for exchange of data and information at regional or national conferences or via the Web.
- Year 2-3: team effort. Facilitate creation of geospatial and ecological data sets by serving as a coordinating entity and a potential funding source.

Note: The above roles could be distributed among the regional NBII nodes to ensure a more thorough and focused means of addressing these difficult, yet important, issues.

Build hierarchical classifications of ecological (functional) attributes of species.

- Year 1: Ongoing, team effort. To be accomplished via workshops, literature review, and written documents presented to key scientists for review.

Identify and catalog data universe

- Year 1: 1 coordinator, 2-3 collaborators, 12 months minimum. Review the status of clearing-house repositories, identify gaps in coverage, and develop alternative approaches to filling those gaps (such as automated resource description).

Standardize interfaces to data and models

- Year 2: 1 coordinator, 2-3 collaborators, 12 months minimum. Undertake data resource classification and analysis (first draft for review).
- Year 2: 1 coordinator (different skills than above), 2-3 collaborators, 12 months minimum. Undertake a similar analysis and classification of the inputs required by modeling tools, so that classes of "automated users" or data consumers are identified and used to create focus (see "Lifemapper" system, based on "Darwin Core" federation schema and the GARP modeling tool for predicting species' distributions).

Establish multiple standard data schemas and ontologies

- Year 1: Ongoing, 1 coordinator, plus cooperators focusing on the various schemas. Facilitate the development of the various standardized schemas, as particular needs or classes of data resources are identified, through workshops and other forums for discussion among interested and knowledgeable parties.
- Year 1: Ongoing, same team as above. Facilitate the process of mapping individual data resources to standardized schemas, through node partnership participation, on-site visits and consultations, and grants of hardware where needed.
- Year 2/3: 6-12 months, same team as above. Develop tools ("Wizards") to assist data holders in the process of mapping their data source schemas to one or more of the standardized schemas, and evaluate the suitability of generalizing them using ontology languages. Also develop tools ("Wizards") to assist researchers in designing the schemas or ontologies for new data sets in such a way that they maximally conform to one or more of the existing standard schemas and/or identify needs for new schemas or extensions of existing standard schemas.
- Year 1: Ongoing. Maintain a network for communication with data holders about modifications to existing standard data schemas and the introduction of new standard data schemas.

Establish standard APIs

- Year 1: Ongoing, 3-9 months, 1 person. Review standards for data access and interoperability.

Establish Registry support for standardization

- Year 1: Ongoing, 6-9 months, 1 person with outside cooperative assistance. Adopt a centralized registry based on a common, well-known standard such as UDDI to remove potentially significant development activity required to build a custom registry.
- Year 1: Ongoing, team. Develop a review team with a repository manager having significant expertise in programming and data management/storage.

Document data quality, provide peer review

- Year 1: Ongoing, 6-12 months, 1 coordinator.

Develop data quality documentation guidelines.

- Year 1: Ongoing, 1 coordinator. Develop an expert review panel and evaluate guidelines.

Provide education on information management and modeling

- Year 1, 12 months, 1-2 people. Collaborate with the USGS Cooperative Research Units or Cooperative Ecosystem Study Units at universities and state agencies to develop informatics courses or programs, or integrate coursework from existing informatics programs at a university.
- Year 1, 12 months, 1-2 people. Develop short courses for state agencies that would be run out of, and in collaboration with, the Coop Units, but involve NBII staff or collaborators.
- Year 1, 12 months, 1-2 people. Establish training workshops/short courses on informatics at key centers.
- Year 1, 12 months, 1-2 people. Mount road shows of key model developers who have excellent "success stories."
- Year 1, 12 months, 1-2 people. Facilitate development of texts on ecoinformatics.
- Year 1, 12 months, 1-2 people. Host NBII meeting focused on educating those external to the NBII on the modeling capabilities associated with the NBII.
- (Effort and duration TBD). Establish an electronic journal associated with the NBII that covers issues of information management, modeling, and information transfer.

Provide a service outlet to develop models and maps operating through distributed biodiversity modeling systems

- Year 1, within 12 months, team. Assess options for computational capabilities and determine which of the below options may work for the NBII:
    - FTP site for download of software
    - Demonstration implementation of algorithms
    - Implementation of a test bracket that allows users to check the validity of their implementation of modeling algorithm on test data
    - Users allowed to run own data on NBII nodes via single user interface
    - Provision of a standard middle-ware Web service that allows users to build their own interfaces to the computational service
    - Provision of a distributed computational service such as Lifemapper or other grid-based system for computational intensive algorithms.
- Year 2: team, with cooperators. Implement one or more of the above options for facilitating data access and modeling (Production system level service).
- Year 1, within 12 months, team. Assess and prioritize list of candidates for activities to be supported and potential algorithms.
- Year 1, 1-6 months, team. Assess target audiences and needs.
- Year 2: Ongoing, technical team. Design appropriate interfaces for each target user group.
- Year 2: Ongoing, technical team. Develop strategy to ensure the interoperability of systems including: standardization of data schema, development of Web services, standardization of application program interfaces, and Web service registration systems such as the UDDI.
- Year 1: Ongoing, technical team. Ensure the infrastructure initially developed by the NBII is adopted by NBII partners.
- Year 1: Ongoing, administrative team. Enable greater involvement with information technology research. This might be achieved though outreach to the scientific community, through hosting of workshops, creation of an electronic journal, proactive involvement in existing funded computational technology projects and ITRs such as SEEK and others.

Implement a clearinghouse for existing models

- Year 1: 3-12 months, administrative and technical

team. Develop strategic plan outlining logistics for initial and full implementation.
- Year 2: Ongoing, administration and technical team. Implement clearinghouse.

Establish a decision support system for selecting and using biodiversity models
- Year 1, 1-3 months, 1 person, followed by peer review. Develop specifications for decision support.
- Year 1, 6-12 months, tech team. Implement decision support system.

Systematically assess the current state of biodiversity modeling thought and practice, inclusive of data and information requirements
- Year 1, 3-6 months, 1 coordinator with team. Assess current state of modeling activities.
- Year 1: Ongoing, 1 coordinator and technical team. Facilitate movement toward mechanistic modeling activities.

Track reuse of biodiversity models
- Year 2: Ongoing, 1 person. Implement procedures for tracking, version stamping and updating models.
- Year 2: Ongoing, 1 person. Track requests for models.

B. Concluding Remarks

This workshop venue gave attendees the freedom to explore the potential for a well developed modeling-support enterprise within the NBII. Attendees agreed that this is needed and would be welcome by the community. However, the group was not constrained in any way by the realities of building such a system and hence laid out a very broad and grand vision which if fully implemented would be a valuable asset to the community. In addition, many steps were identified that would launch the NBII on its journey to meet the vision. Thus in the early stages of this program the NBII is encouraged to address two key goals identified from the workshop. First, further engage the modeling community, that is bring the results of the workshop to other venues and engage a far broader audience in the dialog. This will help to attain buy-in and also to better frame the issues. Second, follow-on workshops should be used to identify discrete and doable projects that can be undertaken by the NBII to accomplish this broader vision in time. This approach will help create technologies and information infrastructure needed to address the NBII's larger goals, in a step-wise organized fashion that can evolve to address community needs.

IV. Appendices

A. Acronyms used

API	Applications Program Interface
AAAS	American Association for the Advancement of Science
BDEI	Biodiversity and Ecosystem Informatics (federal inter-agency initiative)
BIOCLIM	(name of a model)
CBD	Convention on Biological Diversity (UN-sponsored international convention)
CBI	Center for Biological Informatics (of the NBII)
DGVM	Dynamic General Vegetation Model
DiGIR	Distributed Generic Information Retrieval (protocol)
EML	Ecological Metadata Language
ESA	Ecological Society of America
FGDC	Federal Geographic Data Committee
FTP	File Transfer Protocol
GAMs	General Algebraic Modeling System
GAP	Gap Analysis Program (program of the U.S. Geological Survey)
GARP	Genetic Algorithm for Rule Production
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System (family of spatial analysis capabilities)
GISP	Global Invasive Species Program (UNEP Program)
GISIN	Global Invasive Species Information Network
IABIN	Inter-Americas Biodiversity Information Network (Summit of the Americas)
ITR	Information Technology Research (National Science Foundation program)
IUCN	World Conservation Union (formerly International Union for the Conservation of Nature)
LTER	Long Term Ecological Research (National Science Foundation program)
MSDS	Multiple standard data schemas (mnemonic device)
NAC-BDEI	North American Consortium for Biodiversity and Ecosystem Informatics
NBII	National Biological Information Infrastructure (federal inter-agency framework)
NISC	National Invasive Species Council (federal inter-agency initiative)
NRC	National Research Council (federal agency)
NSF	National Science Foundation
ODE	Ordinary Differential Equations
OGC: WMS	The Open GIS Consortium Web Mapping Service
OFWIM	Organization of Fish and Wildlife Information Managers
OWL	Web Ontology Language
PBIN	Pacific Basin Information Node (of the NBII)
PCAST	Presidents Committee of Advisors on Science and Technology
PFT	Plant Functional Type
RDF	Resource Description Framework (protocol)
SEEK	Science Environment for Ecological Knowledge (protocol)
SPIRE	Semantic Prototypes in Research Ecoinformatics (protocol)
UDDI	Universal Description, Discovery and Integration (protocol)
USGS	United States Geological Survey
WDC	World Data Center (for Biodiversity and Terrestrial Ecosystems)
XML	Extensible Markup Language (W3C protocol)

B. Workshop participants

1. Contributors in attendance

Blum	Stan	sblum@CalAcademy.Org	415-750-7145	CAL Acad Sci
Crosier	Catherine	crosier@nrel.colostate.edu	970-491-5630	USGS
Davies	Rhyn	rdavies@usgs.gov	808-984-3718	USGS/PBIN
Dorfman	Dan	dorfman@hawaii.edu	808-956-6616	Hawaii Natural Heritage Program
Frame	Mike	mike_frame@usgs.gov	703-648-4164	USGS-CBI
Fornwall	Mark	mark_fornwall@usgs.gov	808-875-2435	USGS/PBIN
Gross	Louis	gross@tiem.utk.edu	865-974-4295	University of Tennessee
Helweg	Dave	david_helweg@usgs.gov	808-956-9588	USGS-BRD-PIERC
Mosesso	John	john_mosesso@usgs.gov	703-648-4079	NBII
Mulligan	Mike	mike_mulligan	303-202-4242	USGS-CBI
Neilson	Ron	neilson@fsl.orst.edu	541-750-7303	USDA Forest Service
Parks	Bradley	bparks@colorado.edu	303-497-6330	UCB-CIRES/NOAA-NGDC
Pyle	Richard	deepreef@bishopmuseum.org	808-848-4115	Bishop Museum
Sowa	Scott	scott_sowa@usgs.gov	573-441-2791	MoRAP, University of Missouri
Stockwell	David	davids@sdsc.edu	858-822-0942	Univ. of California, San Diego
Sugimura	Tak	tak@mhpcc.edu	808-879-5077	MHPCC
Turner	Robb	rsturner@utk.edu	865-974-4585	SAIN, ORNL
Vieglais	Dave	vieglais@ku.edu	785-864-7792	University of Kansas
White	Denis	white.denis@epa.gov	541-754-4476	US EPA
Willet	Karen	kbwillett@ucdavis.edu	530-752-5678	UC Davis / CAIN

2. Contributors not in attendance

Kineman	John	John.j.kineman@noaa.gov	303-497-6900	NOAA-NGDC
Quinn	James	jfquinn@ucdavis.edu	530-752-8027	Univ. of California, Davis

C. Workshop prospectus

Biodiversity Modeling Workshop, Maui, April 2003

Goal

Convene a working meeting of researchers and practitioners concerned with predicting the occurrences of ecologically important biological elements in order to assess related opportunities and constraints and to shape a research agenda, establish pilot applications, and develop a vision for the role of the NBII in such future activity.

Purpose

The purpose of this workshop is to provide a forum and a venue for scientists to discuss current biodiversity modeling challenges and potential role of the NBII to science support, data and information resources, computational infrastructure, and to enhance organizational capabilities. This first year workshop will provide valuable input to the NBII and will test whether such a workshop should become a regular activity of the NBII (see following detailed statement of workshop problem and focus).

Background-NBII

- The NBII is the nation’s information infrastructure for biodiversity information.
- One key goal of the NBII is to promote data use and development of applications.
- The time is right for the NBII to begin bridging key data sets and their applications.

Motivating elements for an NBII biodiversity modeling initiative

- The NBII should establish a venue for dialog with the biodiversity modeling community and a way for enlisting that community in its development (particularly models and related data tools).
- As the NBII continues to develop, it will be important for the NBII to demonstrate its contribution to modeling (the theme of this workshop) and to decision support.
- There is a convergence of several agency interests to bring a group together to discuss modeling, supercomputing, and the NBII.

Objectives

- A strategy for integrating needs of the biodiversity modeling community into the NBII (includes program direction, community data needs, and dialog on integration of models into the NBII architecture),
- Identification of supportive projects that should be undertaken within the NBII or in partnership with NSF, NASA, or other interested partners, and
- Recommendations on the future of biodiversity modeling and NBII potential contributions to modeling initiatives, based upon the discussions of the meeting.

Outcomes needed

- A model-based biodiversity research framework for the NBII,
- One-to-several pilot projects under this framework, and
- A strategic and short-term focus for the NBII.

D. Workshop program

Sunday – July 27, 2003

8:00 - 9:00 pm	Evening social hour (travel schedules permitting)
	Goal – Introductions / acquaintance

Day 1, Monday – July 28

6:30 – 7:00 am	Conveners meeting (hosts only, as needed)
7:00 – 8:00 am	Breakfast, group (please join us)
8:00 – 8:30 am	Preview
	(15’) Welcome & charge (Mark Fornwall, CBI)
	(25’) Self-introductions / profiles (participants)
	(10’) Purpose & outcomes needed (Brad Parks, UC-Boulder)
	(10’) Comments and revisions (participants)(limit relative to next item below)
9:00 – 11:30 am	Reconnaissance: “map” the BD modeling universe and identify domain(s) of interest
	(Identify and structure points of discussion to follow, first pass)
10:00 – 10:15 am	Break
11:30 am – 1:00 pm	Lunch
1:00 – 3:00 pm	Informal presentations of modeling/mapping applications by participants (TBD)
3:00 – 3:15 pm	Break
3:15 – 5:30 pm	Informal presentations of modeling/mapping applications by participants (TBD)
5:30 pm	Adjourn for dinner
6:00 pm	Dinner, group (location TBD)
7:00 – 7:30 pm	Conveners meeting (hosts only)

Day 2, Tuesday – July 29

6:30 – 7:00 am	Conveners meeting (hosts only, as needed)
7:00 – 8:00 am	Breakfast, group (please join us)
8:00 – 8:15 am	Preview & recap progress

Day 2, Tuesday – July 29 (continued)

8:15 – 10:00 am	Discussion: state of theory and practice supporting or limiting biodiversity modeling
10:00 – 10:15 am	Break
10:15 – 11:30 am	Reconsider: "map" of BD modeling universe and domain(s) of interest (Identify and structure points of discussion to follow, second pass)
11:30 am – 1:00 pm	Lunch
1:00 – 3:00 pm	Discussion: evaluate and rank alternate modeling approaches
3:00 – 3:15 pm	Break
3:15 – 4:00 pm	Discussion: CBI/NBII modeling mission/vision opportunities and constraints
4:00 – 5:00 pm	Discussion: other institutional modeling missions/visions, linkages
5:00 – 5:30 pm	Review progress / results
5:30 pm	Adjourn for dinner
6:00 pm	Dinner, group (location TBD)
7:00 – 7:30 pm	Conveners meeting (hosts only)
7:30 – 8:30 pm	Special interest meetings (as needed)

Day 3, Wednesday – July 30

6:30 – 7:00 am	Conveners meeting (hosts only, as needed)
7:00 – 8:00 am	Breakfast, group (please join us)
8:00 – 8:15 am	Preview & recap progress
8:15 – 10:00 am	Discussion: matching opportunities to constraints for CBI/NBII modeling support
10:00 – 10:15 am	Break
10:15 – 11:30 am	Cont. Discussion: matching opportunities to constraints for CBI/NBII modeling support
11:30 am – 1:00 pm	Lunch
1:00 – 3:00 pm	Recommendations: action plan for CBI/NBII biodiversity modeling support
3:00 – 3:15 pm	Break
3:15 – 5:00 pm	Recommendations: action plan for CBI/NBII biodiversity modeling support
5:00 – 5:30 pm	Assign completion tasks, follow-up actions
5:30 pm	Adjourn for dinner / departures
6:30 pm	Dinner, group (travel schedules permitting, location TBD)
7:00 – 7:30 pm	Conveners meeting (hosts only, travel schedules permitting)

Day 4, Thursday – July 31

7:00 – 8:00 am	Breakfast, working (conveners/hosts only)
8:00 – 9:00 am	Review key results, consolidate notes, reallocate remaining time
9:00 – 10:00 am	Fit presentation framework to results (Outline, schematics, etc.)
10:00 – 12:00 noon	Map results into presentation framework
12:00 – 1:00 pm	Lunch, working (conveners/hosts only)
1:00 – 2:00 pm	Map results into presentation framework (cont.)
2:00 – 3:00 pm	Assign completion tasks, follow-up actions
3:00 pm	Adjourn