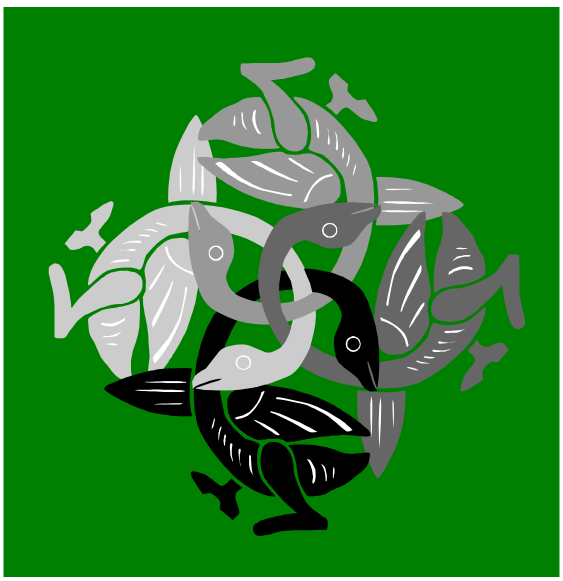
METADIVERSITY II:

Assessing the Information Requirements of the Biodiversity Community

Summary of Findings



Editors:

Barbara Bauldock
Wendy Wicks
Jill O'Neill

Sponsored by the

U.S. Geological Survey, Biological Informatics Office
and the
National Biological Information Infrastructure

and the

National Federation of Abstracting & Information Services (NFAIS)

June 25-26, 2001
Charleston, South Carolina

National Federation of Abstracting and Information Services (NFAIS)
1518 Walnut Street, Suite 307
Philadelphia, PA 19102
Fax: 215.893.1564
Phone: 215.893.1561

Email: nfais@nfais.org

Web address: http://www.nfais.org

NFAIS is a registered trademark.

National Biological Information Infrastructure National Program Office
U.S. Geological Survey, Biological Informatics Office
12201 Sunrise Valley Dr
Reston VA 20192
Fax: 703.648.4224
Phone: 703.648.4090

E-mail: nbii@nbii.gov

Web address: http://www.nbii.gov

NFAIS gratefully acknowledges the support and collaboration of the U.S. Geological Survey, Biological Resources Division, under Cooperative Agreement # 00HQAG0045.

TABLE OF CONTENTS

PREFACE

MetaDiversity II was a symposium jointly organized and sponsored by the National Federation of Abstracting & Information Services and the U.S. Geological Survey. It was held June 25-26, 2001, in Charleston, SC. A previous meeting, MetaDiversity, had been held in Natural Bridge, VA, in November, 1998. It was in response to recommendations from this first meeting that MetaDiversity II was convened.

Specifically, MetaDiversity II was organized to identify the following:

- Places biodiversity information user needs are being met and how they are being met;
- Gaps in biodiversity information needed to support users;
- Populations of priority users of biodiversity information

The meeting attracted an international body of participants from the public and private sector. The group of sixty participants included researchers, educators, information professionals, software providers, and program managers who worked together over the course of two days to identify specific information needs and behaviors of this interdisciplinary community.

The program (see Appendix A) consisted of presentations from fourteen key individuals involved with creation and development of information applications and systems as well as the information seeking behaviors of scientists and researchers who use those systems. Once heard, presenters' ideas and conclusions were critiqued by a peer review panel of individuals active in the field and the floor was opened to general discussion by all meeting participants.

The participants (see Appendix B) were divided into working groups to analyze and discuss user needs, data collection and usage, the tasks facing them in the field and the desirable resources needed to successfully complete work in the field. Meeting participants were asked to hypothesize about the ideal knowledge environment (i.e., a wish list of system resources and functionalities) that would enable scientists and researchers to accomplish their aims and meet specific research objectives. From these active discussions emerged a list of recommendations and courses of action necessary to improve access to biodiversity information on a global scale. The recommended actions are appropriate for further development by the National Biological Information Infrastructure in particular and by the biodiversity community in general.

The following report provides a summary of the discussions and recommendations that were formed at MetaDiversity II.

INTRODUCTION

Background

On November 9-12, 1998, in Natural Bridge, Virginia, a international group of biodiversity researchers met with a group of experts in metadata creation and management to exchange views from their respective areas of expertise and to discuss the application of metadata to the discovery and management of biodiversity information. That symposium, entitled ***MetaDiversity***, was jointly sponsored by the Biological Resources Division of the U.S. Geological Survey (USGS) and the National Federation of Abstracting and Information Services (NFAIS). The symposium was designed to respond to the national and international grand challenges in biodiversity information management by helping to define how metadata could contribute to, support, and enhance the biodiversity research agenda through its incorporation into an information infrastructure to support scientific advances.

This landmark meeting — attended by 82 representatives from government agencies, academic institutions, nonprofit non-governmental organizations, associations, national laboratories, funding agencies, and commercial firms from the United States and abroad — began an important interdisciplinary dialogue. The participants ended their meeting with a call for community in biodiversity.

MetaDiversity Redux

In response to this call for community, ***MetaDiversity II: Assessing the Information Requirements of the Biodiversity Community***, was held June 25-26, 2001, in Charleston, South Carolina. Again jointly sponsored by USGS and NFAIS, the meeting provided approximately 60 members of the biodiversity community with a forum to discuss and define community needs and requirements. As with the first MetaDiversity meeting, participants came from a variety of sectors, public and private, with interest in biodiversity information: federal agencies, publishers, abstracting and information services, libraries, universities, and research institutions. The group included researchers, educators, librarians, information specialists, and program managers.

As a direct follow-on to issues raised in the first MetaDiversity meeting, the program for MetaDiversity II was designed to identify:

- Places where the information users' needs were being met and how they were being met
- Gaps in biodiversity information needed to support users
- Priority populations of potential users

To those ends, MetaDiversity II featured authorities on user needs and user behaviors who shared the platform with speakers describing various efforts to develop key information resources, from geographic information systems to databases of the primary literature to a variety of other types of information resources. Several publishers addressed model for access to information, and a biologist's biodiversity informatics wish-list invoked an over-arching vision for the group's deliberations.

To stimulate further dialogue, a peer review panel — Jeff Waldon from the Conservation Management Institute at Virginia Tech, John Porter from the University of Virginia, and Crispen Wilson from Conservation International — challenged the presenters on how they built their information resources, how they identified user needs, and why they feel what they have done has fulfilled a user need.

Scientists read an average of 130 scholarly articles per year (up from 100 articles in 1977). They read from an increasing number of journals each year: 25 in 2001, compared to 18 in 1995. Approximately 50% of the readings contain information new to the reader. Journals are considered important compared with other resources.

While exchanging views and sharing information among the diverse participants was key to the seminar, MetaDiversity II was primarily a working meeting. After listening to and querying the presenters, who established the context for the meeting, the attendees addressed themselves to informatics solutions for the biodiversity community. Four working groups, comprising both attendees and speakers, shared their expertise and understanding to enumerate the needs and define informatics objectives for the biodiversity community. From these working groups, which met three times over the course of two days, came a list of recommendations concerning priorities for the community in general and for the development of the National Biological Information Infrastructure (NBII) in particular. NBII (http://www.nbii.gov) is a broad collaborative program to provide increased access to data and information on the nation's biological resources. NBII links diverse, high-quality biological databases, information products, and analytical tools maintained by NBII partners and other contributors in government agencies, academic institutions, non-government organizations, and private industry.

Future Plans

The dialogue begun at MetaDiversity and continued at MetaDiversity II has by no means concluded. Presenters and attendees alike agreed on the desirability of continuing to building the community and to convene again to assess progress on the MetaDiversity "wish-lists" and recommendations arising from MetaDiversity II. Plans are already underway for ***MetaDiversity III*** in 2002.

ASSESSING COMMUNITY NEEDS

Experts in the assessment of community information needs provided background information to MetaDiversity II participants on the information requirements of both researchers in general and the biodiversity community in particular.

What Do We Know About Scientists' Use of Information in General?

Don King, University of Pittsburgh, and **Carol Tenopir**, University of Tennessee, authors of *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*, shared the seven lessons learned from a national survey of 14,000 scientists in all fields of science, in both university and non-university settings, and representing over 100 organizations.

Lesson 1: Scientists use multiple means of communication.

Written communications include personal correspondence, preliminary findings, formal progress reports, patents, convention presentations, manuscripts, final technical reports, theses, preprints, journal publications, books, reprints, databases, and abstracting and indexing publications, among others. Oral communications include informal discussions, local colloquia, special group meetings, information conferences, and state, regional and national conferences, as well as simple conversations among colleagues. Annually, scientists spend an average of 166 hours in information discussions, 290 hours reading, 98 hours preparing and making presentations, and 175 hours consulting or giving advice.

Lesson 2: More scientists mean more literature.

The growth of journal literature is correlated with the number of scientists, with one article per ten scientists generated yearly.

Lesson 3: Scientists read a lot and find reading essential.

Scientists read an average of 130 scholarly articles per year (up from 100 articles in 1977). They read from an increasing number of journals each year: 25 in 2001, compared to 18 in 1995. Approximately 50% of the readings contain information new to the reader. Journals are considered important compared with other resources.

Lesson 4: Readers are price sensitive.

Journal prices have risen 6% per year since 1960. The average journal subscription is now $1,100/year. The average number of personal journal subscriptions per year per scientist fell from 6 in 1977 to 2 currently.

Lesson 5: Scientists use a variety of ways to get journal articles.

Sources include library subscriptions, inter-library loan, reprints, unbundled journals, shared department or unit collections, and colleagues.

Lesson 6: Separate copies are becoming more prevalent.

As of May 2001, the Los Alamos National Laboratory preprint archive of high-energy physics papers (http://xxx.lanl.gov) was averaging 200,000 connections per day; 35,000 new papers are expected to be added to the preprint archive in 2001. Each article gets an average of 300 downloads per year. Electronic preprints accounted for 3.6% of all reading for Oak Ridge National Laboratory (ORNL) scientists. The number of searches of PubMed, a gateway to health information (http://www4.ncbi.nlm.nih.gov /PubMed), ranges from 400,000 to over 1 million per day. Scientists are now reading more titles and more separate copies than previously because readings are identified separately by online searches and by their colleagues.

Lesson 7: Electronic journals are adopted when it is easier.

In 2000, 35% of ORNL scientists' readings were from electronic sources; over half of these readings were accessed through library electronic subscriptions, free web sites and personal electronic subscriptions. One-fifth of medical faculty article readings were from electronic sources. The amount of time spent reading electronic articles was similar to paper-based articles, although identifying and locating electronic resources takes more time than print. The disciplines experiencing high use of e-journals are high-energy physics, medical sciences, and biological sciences, and electronic journal use is increasing in general. Peer review of electronic sources is important to many. Non-core readers are price sensitive, and students prefer electronic journals.

The data from the study do include the life sciences, including the biological sciences, as defined by the National Science Foundation (NSF), but the data have not yet been broken out by the actual amount of reading, etc. King and Tenopir have focused more on understanding the range and extremes. Medical faculty comprise one end of the spectrum; that group tends to read more, rely on a fewer number of highly-regarded titles, read more articles from each of those, and rely more on journal articles. At the opposite end are engineers, who tend to rely on many different types of information sources. Engineers are very highly reliant on oral communication channels and technical reports and less on published journal literature, but as a group spend more time reading a particular journal article when they find one of interest and relevance. Other fields, including the life sciences, are in the middle of these extremes.

Numerous studies, particularly those published by scholarly societies and publishers, discuss what attributes an electronic journal should add over those available through a print edition and what value is added by e-journals. The research clearly shows that people want e-journal attributes that are different from the print journal, including links to data and the ability to manipulate data.

> ***It appears that information is becoming more fragmented. Even when a "public library of science" has been created, scientists will not want to search the entire library; rather, they will want information relevant to them — based on user profiles, for example — delivered automatically.***

Yet right now most scientists, in almost every field, are using journal literature in electronic form very much the way they use journal literature in print form. They look at it, print it out, read it, and digest it. The one difference in use seems to be when electronic publications present articles in a disaggregated format, e.g., articles and other material from various sources grouped by general or specific subject matter (as opposed to the more traditional aggregated model, where articles are presented as a part of a single journal issue). It is perhaps too early to determine how much these usage patterns have to do with habit and how much is inherent in a reader's expectation concerning referred, published articles. Habits may change as models for electronic journals become more diverse.

Toward Comprehensive or Segmented Information Systems?

It appears that information is becoming more fragmented. Even when a "public library of science" has been created, scientists will not want to search the entire library; rather, they will want information relevant to them — based on user profiles, for example — delivered automatically…which is essentially what journals do now. Even if scientists choose to browse more comprehensively, they will ultimately return to a more pragmatic and focused manner of looking at things: discipline-based and segmented. All of the world's scientific information in one big chunk will become more and more overwhelming to researchers and less and less the way people will want to do science.

Portals and gateways are important value-adding features of electronic delivery. These types of guides can identify the highest quality information for researchers' particular fields of science and lead them to information of greatest value to them. We can build a large information collection with multiple doorways to it, but this goes against the client-server computing models of today's world. We want more specialization, but that poses challenges to interdisciplinary research. On the other hand, as interdisciplinary research increases, more fields of study emerge, which leads to the publication of journals addressing the needs of those new, specialized fields.

As these new fields of study develop, a core collection of journals and documents will be important to those specialties. Large databases are useful to provide access to information available outside that concentrated in a small number of journals.

The Cost of Information Access

Access is currently thought of in terms that are as much economic as intellectual. A good deal of attention has been paid to the costs associated with access to information. The research community, represented by both information professionals and scholars, is challenging the current structure of scholarly publishing and the business models on which it operates, even to the point of boycotting publications that are deemed exorbitantly expensive. Content providers, scholarly associations and commercial publishers support a variety of formats and technologies for delivery of content but know that the fixed costs of delivering content have not diminished, even as page charges, subscription fees, and licensing models are re-examined and revised to accommodate the needs of the customer base.

What Do We Know About Scientists' Use of Numeric and Other Kinds of Data?

Paul Uhlir, of the National Academy of Science, reported on the findings of a 1996 National Research Council study entitled *Bits of Power: Issues in Global Access to Scientific Data*. The study was organized by the U.S. National Committee for CODATA to examine the then-current state of global access to scientific data; identify strengths, problems, and challenges; and recommend actions that build on these strengths and eliminate or avoid identified problems. The report is available on the National Academy Press web site (http://www.nap.edu).

Bits of Power identified the following information technology trends:

- Decreasing costs of computing and communication
- Enhanced capabilities for collecting and analyzing scientific data
- Advent of digital wireless
- Increasing exploitation of broadband network.
- Shift from public to private dominance on the Internet
- Increasing technological capabilities for supporting scientific collaborations
- Growing capabilities for natural language processing — machine translation of voice and text
- Increasing recognition of the importance of standards

- Greater cooperation in monitoring and controlling network activity
- Increasing use of intranets

The technological issues and concerns related to these trends are:

- Inadequate description and indexing of data — due to the idiosyncratic nature of research
- Rapid obsolescence of electronic media — nothing lasts more than ten years
- Technological inadequacies and high cost of access in developing countries

In parallel with the information technology trends are the scientific data trends:

- Rapid growth of the body of scientific data
- Development of large international research programs
- Insufficient funding for data management and distribution
- Decentralization of data management and distribution
- Increasing recognition of needs and data management challenges

A major issue is the lack of compliance with the policy of full and open data availability in scientific data management. This policy guarantees that data from publicly funded research are made available with as few restrictions as possible, on a nondiscriminatory basis, for no more than the cost of reproduction and distribution.

The implementation of the policy has been less than complete. The U.S. government is the best adherent to this policy; other countries are less compliant. The greatest challenge to the implementation of this policy is in developing countries where the presumption about publicly funded data is that they are classified. Even with its respectable track record, the United States has regressed in implementing the policy in recent years, and some types of data have been increasingly restricted.

While access to and management of all types of scientific data is quite good, certain characteristics of biodiversity research make both access and good management particularly difficult. These include: (a) the extreme heterogeneity of the subject matter; (b) the highly distributed and individual investigator-driven nature of the research; (c) the fact that most biodiversity resources are located in the world's poorest countries; and (d) the growing political, economic, and legal restrictions on data access in both the developing and developed world.

Can the Scientific Community Influence Policy?

At best, said Uhlir, scientists are the tail wagging the dog. Generally speaking, it is difficult for the scientific community to influence the legislative process or government policies related to information management and access.

A good lesson can be learned by noting the differences between the U.S. and the European Union (EU) situations relating to the legal protection of databases. In the EU, a directive (http://europa.eu.int/eur-lex/en/lif/dat/1996/en_396L0009.html) was enacted in 1996 without any real consultation with any of the public interest constituencies in the science education or library communities. In the context of intellectual property law, we typically want a balance between the rights-holders and the users. There is certainly always room for reasonable debate on where the balance should be. If the scientific community is not at the table to present its views and interests, then it will not be heard.

> ***The National Science Academy's position on database protection has been to support legislation in the United States because, although there are interest on the other side of the issue, we see a gap in the law that does require additional statutory protection for non-copyrightable databases.***

The process to amend intellectual property law, both at the international level (through the World Intellectual Property Organization, WIPO) and in national legislatures, is typically dominated by high-paid lobbyists for the large internationals (like Microsoft) and large publishers (like AOL Time Warner). These are the people who want maximum protection for their information, and they have little regard for science or public access. The playing field is very skewed. That is yet another reason for the scientific community to get involved.

In the United States, the debate was much broader. It drew major attention in Washington, although database protection legislation was clearly going through without any debate initially. This mobilized the science community, as well as the U.S. Office of Science and Technology Policy and the Office of Management and Budget, and got a larger set of players involved. That put a halt on the steamroller that was happening at the time. Subsequently, the scientific community — the educational and library communities — got involved in the legislative debate, and other large economic concerns (the Baby Bells, the Chamber of Commerce, Charles Schwab, the New York Stock Exchange) have been involved in this, on both sides of the issue. There has been generally an impasse in legislation in the United States, yet there is a reasonable chance that any ultimate legislation will be substantially more balanced than the EU version.

It is important to encourage colleagues in developing countries to get involved in the legislative process through their ministries and intellectual property offices, or to speak to the ministries that deal with the biodiversity legislation and try to enter more balance into that equation with regard to distinguishing better between basic research and commercial exploitation of biodiversity. It is to the advantage of those countries to encourage basic research. As it stands, it is not a well-balanced regime.

The National Science Academy's position on database protection has been to support legislation in the United States because, although there are interests on the other side of the issue, we see a gap in the law that does require additional statutory protection for non-copyrightable databases. There is also the issue of the EU directive being the only legal model internationally, which is, in our view, very negative. It would be better to have a U.S. model that is reasonable as a countervailing approach.

Unfortunately the debate in Congress has gotten extremely politicized and polarized. The people in the middle who would like to see some reasonable legislation are being out-shouted by the two extremes. To the credit of those who are pushing for database protection, the more protectionist version in the House Judiciary Committee has moved much nearer to the middle.

What Do We Know About Scientists' Perceptions and Uses of Electronic Resources?

Narrowing the focus slightly, **Sharolyn Aschenbrenner** of JSTOR discussed the outcome of an extensive survey conducted by JSTOR in September 2001 to measure faculty perception and uses of e-resources. JSTOR (http://www.jstor.org) is a nonprofit organization seeking to build a reliable and comprehensive archive of important scholarly journal literature.

On the basis of a grant from the Mellon foundation, JSTOR retained a commercial market research firm, Odyssey (San Francisco, CA), to embark on a study of how U.S. academics currently perceive and use electronic research resources in general. The objectives were to understand the academics' attitudes about the current and future impact of technology on research and teaching and to gain insight into their awareness and attitude toward archiving and JSTOR. While this particular study focused on social science and humanities faculty, the attitudes revealed had applicability to this audience.

A sample of academics was chosen by random selection from a list of over 150,000 faculty members. A total of 4,220 professors at U.S. colleges and universities completed the surveys.

The results of the study revealed that more than 60% of faculty:

- Are comfortable using electronic resources
- Believe a variety of resources are important to their research
- Consider electronic databases to be invaluable
- Believe they will become increasingly dependent on electronic resources in the future
- Currently use online catalogs, full-text electronic journal databases, and abstracting and indexing databases

JSTOR concluded that:

- Electronic resources are important to faculty
- Humanists depend more on the library for access than do social scientists
- The role of the library in information access is expected to diminish
- Electronic archiving is important to all

THE VISION

A Biologist's Informatics Wish-List

In breaking new ground, it is useful to take a critical look at past and present metadiversity models. **Meredith Lane**, of the Academy of Natural Sciences, presented a vision for biological informatics in her dinner speech. Dr. Lane traced a long history of linear information transmission. She remarked, "The mark-up languages and the hyperlinks give us the freedom to be reticulate in our knowledge. The problem is that we have not truly accepted this freedom and accepted the responsibility that goes with it."

If you ask someone to build an information system that will help you in your work, Lane stated, what the systems person will automatically design is an "*a priori*" model. Such a system asks the question first and then figures out how to answer it and how to get the machinery to return an answer. This is typical of The Nature Conservancy (TNC) model of database building, and it is extremely useful, as well as expensive. The centralization of the database maintenance is in some ways more efficient and in some ways more expensive. An enormous expense is the human time and effort put into essentially feeding the machinery. Nonetheless, it's valuable and may be the way to go in certain circumstances.

But for the larger purpose of really turning the Gutenbergian kind of information system into something dynamic and useful for questions we cannot presuppose, we have to take a different tack, suggested Lane.

Lane outlined the problems with the *a priori* model. If the data schema are set up to answer certain questions and then someone wants to add another question, the designer has to go back in and rearrange and adjust the schema. And this implies going back usually to the data sources to get more data. The majority of the human effort is in feeding the machinery. The human intellectual power has been used to massage data before it is actually put into the dynamic process.

Contrast that with an *a posteriori* model where you can have lots of different types of queries asked against the same type of databases. This will be a hodgepodge. Different groups, at different times, for different reasons, will be putting together many small databases that will be locally controlled, yet can be useful in a global sense. The place where the human effort and intellect goes in will be much more appropriate for the use of our brains. The intellect is engaged at the analytical and synthetic level rather than in feeding data into the machine.

If a new question is added in a system like this, there is no call to rearrange the schema automatically. Instead the designer develops a new analytical, synthetic, or query tool. And maintenance is distributed, which means that most of it will be there most of the time. This model is truer to a knowledge base than an infobase. However, it does require metadata standards.

An example of a knowledge-based system that is emerging along these lines is Neodat. There are about 70 institutions involved in Neodat. The developers of this system did their standardization by common hardware and software, rather than employing metadata standardization. But there are limits to the ability to do that, and indeed, it would be hard to push much farther in the current direction that Neodat is going; the metadata is more important.

What do we need to do? Individuals who are providing data need to learn more about how they know what they know, and they need to adopt mechanisms that are expandable and more flexible than is the tendency in original research. If a scientist collects five data points for each site visited, and then adds one more data point,

> ***Individuals who are providing data need to learn more about how they know what they know, and they need to adopt mechanisms that are expandable and more flexible than is the tendency in original research.***

the scientist could add validity for someone else down the road. This is the desired approach. Our culture needs to change to reward these types of efforts.

At the top of Lane's informatics wish list are:

- Computational capacity — query languages, mechanisms for getting in and out of various databases with and without metadata, hardware, and software capacities
- Connectivity — bandwidth, the actual connections into places that do not currently have them
- Content — each and every little database should somehow be accessible

Lane's IT wish-list adds the following items:

- Data entry tools: ways to get more data into the system — keyboard, voice recognition, video, scanner, direct digital feed, and even more innovative means
- New data from research and monitoring (includes specimen collection data): direct capture from GPS, other instrumentation, immediate entry by researcher, etc.
- Data from static media (legacy data, books, journals, card files, specimen labels): OCR, VR, natural language parsing, quality control mechanisms, etc.
- Software tools: Automatic data description — tools to identify fields automatically; data cleansing, data conversion, indexing and cross-database linking tools

Lane commented that some of these initiatives have to be executed at the global level, and that is what the Global Biodiversity Information Facility (GBIF) (http://www.gbif.org) is about. GBIF is needed to establish a worldwide resource for geographic and taxonomic names to carry biodiversity informatics forward.

Lane's comments formed the basis of a lively dinner discussion. How could the diverse inter-disciplinary nature of this particular research community best be served in realizing her vision? The audience was drawn to respond.

Question: I agree about the data entry stuff as being extremely important. How do we get it done?

Answer: Some of it by dint of the hard work of the undergrads. You enter the data at the most atomic level that you can, and then you can build all kinds of collections from that.

Question: Talking about citizen scientists. How do we get citizen input into these databases?

Answer: Last spring, I was associated with a project called Birdcast — using radar to track birds as they migrated — and there was an area for citizens to add data. There is probably a vetting process to be done there. There is also a doubt about the quality of the scientific data submitted by laymen. On the other hand, scientists may be as doubtable as citizens are.

Comment: There are two good examples of documenting the evidence to back the information we use: 1) an effort to ascertain solid waste pollution in the marine environment in this country and around the world, an effort accomplished by people going out and collecting all of the data from several places around the world every year; and 2) REEF — a private group gathering information on tropical fishes — that trains amateurs to go out to the same locales year after year and identify tropical fishes, and then, through a roving technique, to do estimates of population.

If we document the information that is in databases, people can ascertain whether they are going to trust it or not, depending on who is providing it and the methodologies that were used.

Comment: What about the GLOBE project that involves school kids in gathering data? I think that is very much along the same lines as the project just referenced. What this does is, on an ongoing basis potentially, engage youngsters who get trained and then participate, and finally see the value of their activities. They can then become engaged in the activity throughout their lives. That is what we need to do for long-term biodiversity conservation in this world.

Adults are also very powerful and important to empower in biology. Volunteers who are big donors often volunteer for Earth Watch or other programs and take data that is supervised by biologists. If biologists want to do good biology, they need money and public backing for that money. Letting the public participate in biology is a very good way to continue that effort.

Comment: I actually believe in specimen-level data, and you need to tag your specimen with a unique identifier. Once you have a unique identifier on it, you can do some amazing things. You can ask a professional systematist to identify something and then, two years later, to identify it again to see if you get the same answer back. And you do not always get the same answer back. Systematists have a tendency to split when you have small samples, and lump when you have large samples.

Question: I was excited to hear you suggest there might be some way of automating the input of label data from legacy (old historic) specimens. Could you elaborate?

Answer: In the digital library world, one of the things we are doing is looking at how to work with legacy publications. Every taxonomic publication ever done usually incorporates at least one specimen, and often a materials study incorporates 10, 15, or even 20 or 30 specimens. We can assume there is a fairly high level of confidence in the published specimen data you see in legacy and current publications. It is possible to extract those data and to verify them — to the specimen level — and we are experimenting with it. Certainly it is possible to do it in prospective publishing — to adopt protocols that make the data immediately available as they are publishing it for biodiversity use.

Question: I wondered if there was a little bit of a false distinction between the *a priori* and *a posteriori* models. I think you can overdo the questions upfront, and I think you want to strip away as much as possible towards the roots of the types of data information. But, how, without agreeing on a sort of standardization that presupposes a certain type of information needed by everyone, would you get landscape, global, or larger views?

Answer: I made the distinction larger than reality. I was trying to get more at the thought of the individual provider asking more questions.

Comment: We have been computerizing probably on the order of six million specimens in ten years, and most of it was done by hand and by human eyes. I think it is a major challenge to do it in other ways. Some form of pattern recognition can be used for the forms of herbaria, though it is very complicated to do it. But for insects? There are two possibilities: involving children, like they did in Namibia, and using voice recognition, as they did in Alberta with high school students. But I do not think there are really high-tech solutions right now.

ADDRESSING COMMUNITY NEEDS

Collaborative Biodiversity Information Systems

The MetaDiversity II program featured a wide range of representatives speaking on various biodiversity information systems currently under development. The presenters were asked to describe the nature of their organizations' work with the biodiversity community and to address the following questions:

- How do you assess user needs in building your systems and applications?
- What is known from your user needs assessment?
- Do you fulfill the user needs that were identified?

Each presentation or set of presentations was followed by a brief Q&A.

RED MUNDIAL DE INFORMACIÓN SOBRE BIODIVERSIDAD

Jorge Soberón, Executive Secretary of the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO, the National Commission for the Knowledge and Use of Biodiversity), explained that CONABIO (http://www.conabio.gob.mx) was created to promote and coordinate knowledge, conservation, and sustainable use of biodiversity of Mexico. Its first objective is to create and maintain an updated national biodiversity information system. Users of the system are government officers, scientists, experts, consultants, agriculturists, foresters, and the general public. These users' needs are highly varied.

The system developers believe "whereabouts" questions are relevant to users — questions such as "Which species are to be found in a given place?" and "Where can I find a given species?" Such questions can be answered using the data in scientific collections. REMIB, the Red Mundial de Información Sobre Biodiversidad (World Network of Biodiversity Information, http://www.conabio.gob.mx/remib/remib.html) developed by CONABIO, is based on specimen data: specimen databases; catalogues and authority files; remote sensing capabilities and electronic cartography; species databases; statistical and analytical tools; and expert networks.

To maintain, update, and analyze the data in these large databases coming from many sources, CONABIO uses two tools, the web search engines in REMIB and *Species Analyst*. "You need analytical tools to assess the completeness and bias of the data and to extrapolate to obtain distributions, richness estimates, etc.," asserted Soberón.

Soberón stated that the lesson learned in developing REMIB has been: "Good metadata, quality control, and extrapolation tools — combined with raw data — can lead a lot of users to the data."

Panelist Q & A —

Question: Do your users actually use your metadata?

Answer: The metadata is available for remote sensing and graphical data, and I think users do use it.

Question: The problem with working with collections of data is that some taxa are much more likely to wind up in museum collections, like plants, for instance. Are you seeing a similar trend in Mexico?

Answer: Yes. Some taxa are very easy to capture, such as plants, birds, mammals; others are not.

Question: In terms of the data collections efforts, starting with the existing data, what do you see as the critical next steps? Is it sufficient to have extrapolation? Or do you have to start to go out collecting new data, and if so, what are the priorities — types of new data, distribution, etc.?

Answer: I'm not satisfied yet with the data from the existing materials in the museums. Less than 5% of the holdings are computerized. It's just the beginning. We would like to continue. Basic species is a good example where you really need everything. The data have been tested to a certain rough scale, but if you go down to much finer scales, then you really need to go into the field and to start collecting the presences of species data, population dynamics, interactions, community structures, etc. But this is very expensive.

Question: To what degree are you making efforts to validate the models that are being used, the extrapolation, etc.?

Answer: These models are documented in the literature in terms of how well they are working for a certain group for a certain researcher. Right now we are in the middle of assigning large-scale statistical tests of the major models we use.

Question: With respect to the normalization of the heterogeneous data, do you see that happening at the institutional level or do you see that happening "on the fly"?

Answer: This is one of the most pressing problems. I don't think a single institution will be able to normalize or standardize things. We need to work on this problem collectively. There are errors in even the most prestigious museum collections in the world. But it is feasible to find errors; it is a matter of organizing.

Question: We can't compute our way out of all information needs and problems. You are fielding complex queries by hand using human assistance, and that is something we continue to believe and assert and move forward on. How are you providing that assistance? You know you're serving about one person/one query a day. Have you thought about who you are not serving who could use that service? Do you track failed searches, or number of failed searches against volume of inquiries?

Answer: We don't have any assessment of the failed queries. We often get feedback from people we provide answers to, but I can't really answer the question of who we are failing. We do put a high premium on experts and the human brain. We use a network of experts from Mexican and American universities. Practically everything is checked by someone with a brain. We have a group in CONABIO that structures the first draft of an answer, but then we go outside and get outside input, *pro bono*, from scientists. Because CONABIO is a granting agency for scientists, we have good working relationships with scientists, and they are happy to help.

THE LONG TERM ECOLOGICAL RESEARCH NETWORK (LTER)

James W. Brunt, University of New Mexico, described the Long Term Ecological Research Network (http://www.ecoinformatics.org) as a collaborative effort involving more than 1,200 scientists and students investigating ecological processes operating at long time scales and over broad spatial scales. Established in 1980 by NSF, the network now consists of 24 sites representing diverse ecosystems and research emphases.

The network's mission is to provide an understanding of general ecological phenomena that occur over long temporal and broad spatial scales, conduct major synthesis and theoretical efforts, provide information for the identification and solution of societal problems, and create a legacy of well-designed and documented long-term experiments and observations for use by future generations.

"LTER's user group was really ourselves," noted Brunt. Strategies used in the development of the network were to build prototypes using a variety of different technologies, evaluate the prototypes for functionality and interoperability, and design a modular framework from the results. A major hurdle to developing a truly interoperable information system is the lack of a rich and flexible metadata model.

LTER is collaborating on a research project with the National Center for Ecological Analysis and Synthesis and the San Diego Supercomputer Center to explore "A Knowledge Network for Biocomplexity." The objective of this collaboration is to create a national network for data sharing that accommodates data heterogeneity and metadata heterogeneity and enables advanced services such as data integration, quality management, hypothesis modeling, visualization, and analysis.

Three products have emerged from LTER's work: (1) Ecological Metadata Language (EML), (2) Metadata Server (Metacat), and (3) Data Management Software for Ecologists (Morpho). Other efforts include building a semantic web for ecology, building information systems as procedural specifications and ensuring integrity in referential databases.

THE INTEGRATED TAXONOMIC INFORMATION SYSTEM (ITIS)

ITIS is a growing international partnership between federal agencies, NGOs, state agencies, and academic cooperators, explained **Janet R. Gomon** of the Smithsonian Institution. ITIS is an evolving standard taxonomic reference system, a dynamic database of biological names and related data, and a portal for taxonomic information. The system is committed to a standards-based infrastructure.

The database covers all kingdoms, but it is a thin database. The data element groups include a unique identifier, accepted scientific names in a working classification hierarchy, synonyms and common names, author/date for name, sources/source publications for names, distribution of species, jurisdictional information (native or introduced), and quality indicators (level of expert review, etc.). ITIS is multi-lingual — English, French, Spanish.

Gomon said, "It's the thin database that provides the data source that's the linking tool to the other databases." It provides a shared "common denominator" for accessing and aggregating biological information. People are downloading the ITIS database (http://www.itis.usda.gov) into their own application and using it as a standard reference source.

ITIS is still working to meet the challenge of biodiversity integration to characterize biodiversity over time and space. There is not a standard classification of organisms that everybody agrees to. There is a need for support for multiple classifications and categorizations and for a taxon concept-based system that is dynamic, as well as for a data exchange capability.

The problem is that multiple names can apply to one taxon concept, and one name can apply to multiple concepts. Names, concepts, and their relationships are dynamic. The solution is a new model for a three-entity relationship — an assertion that represents a unique combination of a name and a reference — circumscribing a taxon concept. Such a solution is being proposed by the Biological Nomenclature and Taxonomy Data Standard Project, an initiative of the Federal Geographic Data Committee.

The Project is in its early stages, undergoing formal public review. It supports the concept that references to different organisms made at different times, in different places, by different authors should be unambiguous. In the future, the standard should provide a consistent reference system for maintaining and cross-referencing dynamic nomenclature and taxonomy in a continuously updated, perfectly archived, and viewable taxonomic database — to enable citation in the literature and a registry function.

Panelist Q & A —

Question: The longevity of federal database projects related to biodiversity has been limited. Do you have strategies in place for the long view of ITIS?

Answer: ITIS started out as a core five-agency Federal effort. This has to evolve into the normal business, normal community infrastructure, of how scientists go about their business.

Question: Considering that ITIS only has two full-time employees, would ITIS be open to receiving taxonomic references from biological organizations for inclusion into the ITIS system, to move towards comprehensive? Is membership open?

Answer: Membership is definitely open. The cooperators are truly building ITIS. Everyone out there needs to build ITIS and make it work. We accept data from many sources. Currently our infrastructure is perhaps a bottleneck in how fast we can do this.

Question: How does ITIS deal with limited degrees of taxonomic resolution? Will ITIS be able to provide names where the user is not able to get down to the species level?

Answer: Perhaps the best area where that comes up relative to Federal agencies' being able to meet their requirements is with the Endangered Species Act, moving at least from the invertebrates to distinct population segments or evolutionary significant units below the subspecies level where a lot of the distinguishing of those units is through molecular sequence data or molecular markers or ecological data. So we have to address that. ITIS currently can link to GENBANK and, at least at the species/subspecies level, get to some molecular sequence data. You have to drop down and add into ITIS names below that species/subspecies level to meet that need.

DISCOVER LIFE.ORG

Working with the All Taxa Biological Inventory has given **John Pickering,** of the University of Georgia, some insights on how the biodiversity community should move forward. He said it's very critical at this point to have a strategic plan about where our data activities should be going.

Pickering is concerned about the user community that is more than just scientists. For instance, there are 10 million visitors to the Smoky Mountains each year, and a lot of those people want to know about biodiversity for education and the pleasure of knowing about biodiversity. How are the information needs of this huge user community being met? "We fail as a community in getting a lot of this information out there to non-experts. Non-experts won't be able to extract this information out of the web because they don't know the scientific names. The gardeners, foresters, school teachers — those are people we need to be addressing in our data systems," he said.

> ***We fail as a community in getting a lot of this information out there to non-experts ... The gardeners, foresters, school teachers — those are people we need to be addressing in our data systems.***

Pickering identified a need to do studies of all the little things (specimens), and not just the larger ones we know. There is a loss of biodiversity because of the failings of the databases. Who is going to fill in the biodiversity information geographically to complete the distribution maps for the species in the United States? Pickering believes it's not going to be the professional taxonomists or the professional scientists. It's going to be the volunteers — the butterfly watchers, the naturalists, and so on.

Pickering continued, "We are going to need schoolchildren in programs doing discovery in each school yard across the continent to document caterpillars of the butterfly, Eastern Tiger Swallowtail. We need to focus on getting new information in by the millions and millions of records, from millions and millions of people." He is very much in favor of species-level databases. "We need to be taking these unique photographs, these unique specimens, these unique observations, and integrating them into a system so we can put the data in," he explains.

Pickering has developed the Discover Life.Org web site (http://www.discoverlife.org) to address many of these issues. The object of Discover Life.Org is to provide a portal for the public to obtain information about living things at the level which is useful to middle school age children on up through adults. Using photos from a matrix-based key in discoverlife.org, users can pick the image that matches their specimen. The system allows people to do searching on images and then get to the information that they can use.

The system invites schoolchildren and others to enter a photo of, for example, the butterfly they saw in their backyard. Users of all levels can interact with the system and then contribute data about what they're seeing and where and when they're seeing it. Users can keep track of all this information using unique identifiers so that specimen records can be accessed again.

Panelist Q & A —

Question: Who are the people who need to be doing the primary work to make things user friendly?

Answer: Undergraduates can do this work. The software has been written so it's very easy. We have had some taxonomic expertise. It's very important that we get information about identification out there.

Question: Can you envision a way to link efforts to develop online keys to various taxa seamlessly?

Answer: I would love to think that we'd have competing keys rather than no keys. The nice thing about these keys is that they are set up to recognize different names and to have links to other data sites. The way it has been designed is to have this cascade of identification guides. We just want to be the portal. We don't need to reinvent the wheel. Other keys are better, and we'll link to them.

Question: How would you identify species that are common vs. incidental?

Answer: The key is presenting you with the characteristics based on how common those characteristics are relative to the rest of the species you have. One of the things you can do once you've linked your identification guides to your databases is to add the phrenology as a key character. You can have sound recordings, as well. The technology is increasingly there.

NATURAL HERITAGE NETWORK

The Association for Biodiversity Information's (ABI) mission is the development, management, and distribution of information about biodiversity, said **Mary Klein**. It oversees an international network of natural heritage programs and conservation data centers — independent programs in all 50 states and some non-state entities (Native American tribes); all the provinces of Canada; and about a dozen Latin American countries and territories. The network supports a system of programs that use common methods and tools to create a larger body of information and knowledge about biodiversity.

ABI (http://www.abi.org) has taken a different approach to developing the next-generation of software by moving the information into a spatial environment. Using a spatial component, ABI plans to build more open systems with more extensibility for users. The Heritage Data Management System (HDMS) will integrate next-generation tabular information with spatial information. NatureServe is the online distribution system for the network. It is evolving with information from users — providing information about the status and location of 50,000 species and natural communities. The network is also moving into the arena of decision support tools.

ABI's structure has feedback mechanisms built into it: member councils; annual meetings; regional conferences; listservs; standing committees; focused working groups aimed at standards; targeted workshops; and user groups that currently meet twice a week to test the data system in development.

ABI has learned from users that you have to be persistent in trying to get the input you need, understand what it means, and develop the systems to use it. The users ABI targets have stiff financial constraints. The system has to be as efficient and convenient to use as possible. It needs to mirror the users' actual business practices. Development of these systems needs to be subsidized.

In Spring 2000, ABI entered into a partnership with ESRI to help ABI better anticipate where spatial technologies are going and develop systems that will evolve as that technology evolves. ABI is also working with other conservation organizations to increase access to the network. For example, it partnered with Conservation International to form InfoNatura — the Latin American version of the online system (http://www.infonatura.org).

[Editor's note: In late 2001, ABI changed its name to Nature Serve. The new web address is http://www.natureserve.org, although the previous URL, http://www.abi.org, also accesses the Nature Serve web site.]

BIOTA

As **Robert Colwell,** University of Connecticut, sees it, specimens are the fundamental particles of biodiversity — only individual organisms, living or in collections, carry biodiversity information. Everything we know about biodiversity ultimately arises from taxonomically and spatially referenced individuals: what they are; where they are (and used to be); how many there are (and used to be); how many species there are in a place; and how many places we may find a specimen.

Colwell and his colleagues have created *Biota* (http//viceroy.ccb.uconn.edu/biota) as a biodiversity data and collections management application for taxonomically and spatially referenced specimen data. It contains no data,

> ***Everything we know about biodiversity ultimately arises from taxonomically and spatially referenced individuals.***

just the tools for efficient entry, maintenance, query, comparison, and export of data that users supply. Colwell is the Project Director and Information Manager; Jack Longino (Evergreen State College) is the Scientific Coordinator. There are 87 systematist collaborators from 50 Costa Rican, North American and European institutions participating, as well.

Together, *Biota's* collaborators do mass collections using various techniques for collecting insects, mites, and spiders. Every specimen is bar coded and individually databased in *Biota*. So far, they have 130,000 specimens of 10,000 species in 580 families prepared, databased, and distributed to specialists. The proportion of species new to science is very high — typically 80% of the species are new to science.

There are about a thousand users of *Biota* in more than 30 countries and 44 U.S. states. They are individual researchers in ecology, systematics, biogeography, conservation biology; biodiversity inventory projects, museums, herbaria, botanical gardens, private collectors/observers, and natural area managers.

What has turned out to be a market niche for *Biota* is an off-the-shelf low-cost ($125) biodiversity data management solution for individuals, projects, and institutions. The pricing has been mainly to cover the costs of distribution and publication, as well as advertising. The royalties are sufficient to cover the continuing cost of development.

Colwell discussed how his team determined user needs and met them. They started out with their own needs in the lab. Features and tools were added as users asked for them. For ecologists and biogeographers, Colwell added a data matrix export for any selection of collecting events (or localities) and species (or higher taxa), and georeferenced specimen records for geographical information system (GIS) input. For museums and herbaria, he added a single-screen entry option for historical specimens, specimen label formatting and printing, and automatic determination history recording. For inventories, he added an image archive for specimens, species, collecting events, and localities; and image comparison tools for rapid identification. For systematists, he added a species synonym system, type specimen links to species records, and NEXUS character matrix export for PAUP & McClade.

Biota Version 2 is now in beta testing, with major new features and a new search engine. In Version 2, there is an optional onboard web server supporting dynamic access for public or secure queries and secure data input.

ETI BIODIVERSITY CENTER

In 1989, UNESCO initiated the creation of an ETI Biodiversity Center (http://www.eti.uva.nl) at the University of Amsterdam to meet the growing demand for taxonomic information and to underpin biodiversity efforts globally, reported **Peter Schalk,** University of Amsterdam. ETI (Expert Center for Taxonomic Identification) is many entities: a research and development organization by and for taxonomists and biodiversity researchers; a new aid/mechanism to document the earth's biological diversity; a knowledge center for biodiversity informatics; and a nonprofit publishing organization to increase taxonomic output.

ETI also aims to be a common gateway between knowledge providers and a broad range of user communities by facilitating an interactive process of data sharing and compilation between specialists in all regions of the world and stimulating an information flow from academia to society.

Two groups comprise ETI's user community:

- Taxonomists and biodiversity specialists. They seek development of ICT instruments for data management; technical assistance to implement ICT tools; and a mechanism for making knowledge available (electronic publishing).
- The users of taxonomic and biodiversity information in science, education, policy, commerce, and society.

ETI employs user assessments for its ICT tool development. An international workshop was held involving different components of the user community to define the list of demands for the desired tool; the list was prioritized; prototype software was sent to a selected user group for extensive testing; final adjustments were made to the software tool; and it was made available for free, for the widest distribution. To maintain the ICT tool, ETI gets continuous feedback from the user group suggesting adaptations and improvements to functionality and capacity.

What did user group A, the taxonomists and information specialists, ask for? A reliable easy-to-use data entry and management system that supports building taxonomic monographs capable of continuous extending and updating of information content, capable of easily exchanging data with colleagues (through the Internet) and between different platforms, flexible enough to cater to personal needs (customizing), and suitable as an electronic publication medium. They also wanted security; tool independence (import/export); freedom of publishing/disseminating when wanted (no copyright restrictions for data use); a subsidized, warranted electronic publication service; offline and online possibilities; speedy handling of finalized content; and some degree of financial return from resulting information products.

What did user group B — users in all sectors — ask for? Reliable easy-to-use systems; offline and online output; reliable high-quality information; data-rich (broad scope) information products; broad application in different fields of use; regular updates of information content; help desk for technical support; and affordable (low) prices.

The result is a multifunctional interactive software package, Linnaeus II, (http://www.eti.uva.nl/Products/Linnaeus.html) that combines taxonomic (multimedia) databases; hierarchies; a literature database; glossary; method section; computer-assisted identification tools; and a GIS in one standard system. Import/export functions warrant communication with other database and information systems.

Panelist Q & A —

Question: This question is for the *Biota* and ETI systems. Is there an opportunity now for the many sites that run the software to be linked together and allow someone somewhere to set up a distributed search capability for searching across databases?

Answer (Colwell): There isn't such a capability yet for *Biota*. We have begun a collaboration with two scientists at the University of Massachusetts in Boston who are interested in building data harvesting machines based on XML to do species pages and also lists of holdings based on a variety of different entry points to databases.

Answer (Schalk): ETI data is available for sharing and connecting to. Users have programming capacity to talk from their side of the record to whoever is on the other side. It is a matter of talking to the web manager.

Question: It seems the data content standards are causing the most taxonomic and geospatial key challenges. Are any of you considering a centralized database for distributed data entry as a model for overcoming some of these various standards?

Answer (Colwell): *Biota* is a product for people to use, and people can use it as they wish. In the new version, people can use it in the way you are suggesting, since data entry is possible with any web browser. If it all works out for GBIF, it will do much of this. Not all the approaches are on the same level. Once the facility is there, there will be a central portal for accessing the data. We need a central depository for names.

Answer (Klein): We have a slightly different take on that in grappling with the data architecture issues for the Natural Heritage Network. If we are looking down the line to expand the ability for input beyond ABI's member programs, performance is the ultimate issue, close behind that being the issue of institutional control over their own information. We are working on a submission of a proposal to the NSF to look at what is going to optimize online performance and institutional comfort and what the right structure is going to be (not presupposing if it will be distributed or centralized).

Answer (Brunt): MetaCat is designed to be distributed from the onset. It also works in stand-alone modes on the desktop. The idea of being able to manage the plethora of standards is XSL [Extensible Stylesheet Language] technology, maintained centrally. Repositories could work with that XSL.

Question: Where are our opportunities for further interaction to build on the synergies of having different people looking in different ways at these systems? Or are we just better off having independent, isolated systems?

Answer (Schalk): If we started ETI again, we would not just start from scratch but look around at other software packages. On the other hand, there is merit in different approaches, partly because they serve different users and have different contributors. There are many good systems. In the end it is the user who should select. I would like to make an electronic toolbox: a compendium of software tools with descriptions. I expect to see convergence on good ideas.

Answer (Brunt): There is technology available that could enable more interaction between these systems, were the cultural barriers dropped. XML is the common currency for information being used over the Internet, and can be used to interconnect databases. We have the possibility for an exchange standard using XSL. The computer science community is working on this problem. The World Wide Web Consortium web site (http://www.w3.org) contains a wealth of this information.

Answer (Klein): The producers of these systems can look for opportunities to collaborate with each other in the future. Speaking for ABI, we do not have an outstanding track record in doing that. Because the technologies are going to support this more easily in the future, the users are going to expect it, ask these kinds of questions and ask developers to find ways to bring these systems closer together.

Answer (Colwell): I just hope no one has in mind to develop one gigantic piece of software. That would be a disaster. Years ago, the *lingua franca* of data exchange was tab-delimited text files. We do not want to integrate, substitute, or compete with each other, but to interface and allow pieces of software to exchange information.

Answer (Brunt): The Open Source Software community is showing the way, with hundreds of thousands of contributors. We, as a community, could do the same. It really works, such as taking XML as a standard, and then sending it to review committees. We have a lot of the tools we need, but we have not integrated. Standardization and integration is the key.

Answer (Soberón): The key is not to integrate the systems, but to enable these systems to talk to each other and to share the data. It is also critical to integrate the quality assurance procedures; otherwise, you get a hodgepodge of quality. This is an important stumbling block that is not technological.

Comment: You can collapse the problem a little bit if you think in terms of linking data and just concentrate on those fields necessary to link databases and data sources, and then work on the connect problem. Again, there are the metadata issues — metadata is an open set. You want the data to be well ordered. I would put my effort on concentrating on those linking criteria. This also speaks to the idea of master databases. There should be some master databases — again, species names is a major one, but we might think of some others, in order to come up with some common codes that we can use to start linking these databases together.

I would like to make a final comment on networking issues. Biodiversity is a civil society issue — biodiversity is important to livelihoods and the well-being of so many people. So when we are thinking of designing our systems, we have to be careful of clientelism and closed systems. Biodiversity is everybody's problem. I am concerned about systems that might give power and control and access to information and data to certain users and exclude others. In designing the institutions for these things, the challenge is to open them up so a broad range of civil society can have equal access to the information, and to be able to participate in biodiversity decision making.

There is a lot of creativity out there. I think if we put our online data, our models and systems online, in an open fashion, a lot of neat things can happen … if we had a way of indexing out there. One example is the Mercury harvester model at Oak Ridge National Laboratory.

Biodiversity is a civil society issue — biodiversity is important to livelihoods and the well-being of so many people. So when we are thinking of designing our systems, we have to be careful of clientelism and closed systems. Biodiversity is everybody's problem.

We need people like Rob [Colwell] and Peter [Schalk] to aggressively collaborate, as well as others who are not here. We do not need a proliferation of more tools, in the sense of things that are doing the same thing. Coming from the library community where we have had some success in setting standards and building tools that permit collaboration and efficiencies, from my perspective, it looks like the scientific community does not have the discipline to do that.

We have been talking about authority systems, such as ITIS, where there are probably two or three different axes that define our information. One is geospatial; one is taxonomic; and the other is chronological. To the degree that we as a community can start building authorities supporting those high-level sources of resource discovery, we will have made a huge step forward. How can we support the development of these systems as we develop our systems? We need to think carefully about what the trusted sources are for biodiversity information. The web itself is not a trusted source. Legacy literature is a relatively trusted source.

When the astronomy community had this same conversation, XML did not exist, so they made the decision to have a common file protocol, which literally catapulted their research as a community. I do not know how astronomy does it, but biology never has. We may have a way in these new linking tools to get past that.

Part of the reason that we have these differently developed systems, some of which are doing the same things as others, is because of the funding sources that produce them. They were set out to accomplish a particular project; they did that, and the tool persists. GBIF has a model funding system, where an international source of funds is directed at making the whole system work for everyone on both sides of the equator.

It is great to link these tools, and to create these tools, but they will also need to evolve. To try to put together a system that works for us now is great, but we also have to try to put together a system that will work five years from now — when the new methods and mechanisms are possible as well.

Comment: I want to underscore the value of some of the facilities that information intermediaries offer: the thesaurus, the gazetteer, etc. There is no one else in the world that looks at it with that worm's eye view. We are willing to do these projects. We have the knowledge, expertise, and "insanity" level to take projects on, where you are wading into very complex semantics and doing very labor-intensive kind of mundane work that I do not think scientists always want to do. However, it is labor-intensive and expensive. To get funding for these projects, we often have to compete with science. If nothing else, maybe this is some consciousness-raising that librarians can be more cost-effective than scientists at getting this done. Let us divide and conquer, and each do what we do best.

Publisher Access Models

ACADEMIC PRESS

Ken Metzner at Academic Press says the web is an enabler of wide participation. All Academic Press/Harcourt Health Science journals published since 1993 — 400 journals in all — are now online, and the company is digitizing, article by article, back to the beginning. Academic Press has had successful models for licensing access in place since 1996. The access model is akin to a virtual library card for access to the whole electronic library. Many subscribers who had not previously subscribed to these journals are now downloading the information.

The web provides access for marginal users and marginal usages of the material at incremental costs. Academic Press is involved in a number of initiatives to provide access to developing countries. It is involved with Cornell University libraries in producing a CD-ROM of horticultural research articles that are distributed to various research stations in developing countries. The research stations are charged a nominal fee.

Recently, Academic announced the "IDEAL Charter for Low-income Countries," a low-cost licensing formula that applies to the whole country, with charges based on per-capita GDP. Only countries that have signed the Berne Copyright Convention are eligible. Academic is looking for sponsors — universities, nonprofit organizations, and funding agencies — for individual countries. It is involved in a discussion with the World Health Organization and its health information network to extend this concept to health science journals and perhaps widen the definition of the countries eligible.

Something the biodiversity community might consider is using a standard identification system, known as the Digital Object Identifier, to identify elements of databases. Identification using DOIs has a greater utility in the long run than a URL. Presently the DOI is used for identifying an electronic version of an article and as a technique for linking references within articles through a service called CrossRef.

BIO ONE

BioOne is a new and collaborative project to put some of the primary journal literature in the biosciences online, explained BioOne's President **Heather Joseph**. It was launched to address the need for cost-effective publishing alternatives to commercial journals and also the demand for electronic journals. The initial product (http://www.bioone.org) is a richly linked database containing 40 journals representing a wide cross-section of biological sciences. Ten additional titles are expected by the end of 2002. All titles are peer reviewed and all have high impact factors in their fields. At launch, the database contained current year plus at least one year of back volumes.

BioOne has plans for growth in its content development: subject-specific subcollections, raising quality levels, increasing focus on high-impact titles, and adding international content (20-40 additional titles by the end of 2002). Also planned are enhanced links to bibliographic collections and related databases — both primary content and secondary sources — and access expanded as widely as possible.

True to a cost-recovery business model, the organization is structured to keep operating expenses low through outsourcing and brokered strategic partnerships (in-kind contributions for technical work and hosting, as well as supporting contributions for marketing and sales, etc.). The model includes a revenue-sharing pool for publishers: 50% of the net sales receipts go to the societies that publish through BioOne. This revenue represents subscription income for those entities. BioOne is exploring implementation of multiple revenue streams such as site sponsorships, advertising, distribution arrangements, and spin-off products/services. BioOne's business model will be modified in response to experience, and it will share data with the library and publishing communities.

SPECIAL OPPORTUNITIES FOR PUBLISHERS

In general terms, probably the principal area of expertise that publishers can bring to the community process is an awareness of the need for market research, specifically who has the need for access to information and the size of the market or markets for specific types of information. There are also opportunities for primary literature and databases to be connected in a much more specific way. Publishers can provide mechanisms for building those connections, for example, establishing a link between a private sector or nonprofit database and a government-sponsored resource. Publishers are probably the most likely to have access to the funding necessary to undertake these types of joint projects to build connections among a wide variety of source materials.

The History Cooperative is creating a nice web space with links to journals, with the goal of an online learning environment for school children (study guides, videos, movies). It is a way to get other resources into a spot where people can use them and might be a way to use static information tools to link to scholarly journals and other resources.

While biodiversity clearly affects every segment and stratum of society, not everyone is aware of the sources for discovering the advances in the field of biodiversity, nor has the resources to access material most relevant to him or her. One of the issues discussed frequently in the publishing community is the correct economic model for scholarly publishing. A common sentiment expressed by the scientific community is the notion that publishers add value but that distribution of material is not part of that value chain. Rather, scientists can themselves distribute material over the web and in electronic form on the networks.

A lot of the civil society issues are mirrored in similar discussions that the publishing community is having internally regarding the value that publishers add to core material, such as peer-review, formatting, tagging, and other activities that make it easier to communicate. These discussions are, in essence, attempts at finding the best business model — the one that will best serve the needs of the greatest number of individuals in a manner that is cost-effective for both the publisher and the user. No such model has yet been developed, and it is likely that no one model exists. It is more likely that we will see movement around a number of alternative models in the future, as civil society becomes increasingly aware of the need for more information generated within the sciences involved with biodiversity and publishers seek to understand, address and balance user needs and capabilities.

BRIDGING THE GAP

Working Group Reports

During MetaDiversity II, four working groups were convened to address the objectives of the workshop as well as to provide opportunities for participants to get to know each other better, building the community (also an objective of MetaDiversity II).

The first day, working groups were asked to:

- Identify areas where biodiversity information needs are being met
- Identify areas where they were NOT being met
- Identify potential user groups, or audiences

Maureen Kelly, independent consultant, reported that key among the concerns and wishes voiced by the group that she worked with was the infrastructure tools needed to manage biodiversity information, which will in turn allow all the other good things to happen. These included tools such as:

- standards
- data dictionaries
- metadata
- authority files
- controlled vocabularies

Coupled with this wish list is the recognition of a need to provide tools to make creating information, and then sharing it, easier. It is also important in working with certain information to understood why the data was collected. In fact, it is a question that should be addressed *before* the data is collected because one of the properties of information is that it could have come from a wide variety of sources with different purposes, with different objectives, with different levels of capabilities.

Another important issue is how to create incentives that encourage people to collect the data, create metadata, and use standards, and to encourage the re-use of information. Key to those incentives are procedures and mechanisms

> ***The data associated with biodiversity are difference from the data we see in some other scientific domains in terms of the variety of information that is being created, the variety of places where it is being created, and the variety of reasons for which it is being created.***

for attribution, accrediting and rights management, as well as for recognizing that data is being created and used.

The final issue is a vision for making the data broadly available, particularly to policy-makers, with indications of the quality of the data and the level of certainty, in terms of its capture. The data associated with biodiversity is different from the data we see in some other scientific domains in terms of the variety of information that is being created, the variety of places where it is being created, and the variety of reasons for which it is being created. Bringing all this together into a functional system of interconnected links to get the full power of that information for re-use and for the discovery of new knowledge requires that we respect those differences and that we create the processes that help to overcome them.

Gail Clement, USGS Florida Caribbean Science Center, listed the different target groups of users suggested by the working group in which she participated. The list is a representative cross-section of biodiversity users:

- policy-makers
- ecologists
- geologists
- systematists
- curators
- foresters
- resource managers
- fisheries people
- consultants
- nongovernmental organizations
- Congress and legislative bodies that fund biodiversity initiatives
- publishers
- land use planners
- educators in both K-12 and post-secondary
- lifelong learners
- public health professionals
- amateur naturalists
- specialists in different scientific fields

The group made an important distinction within those user groups between very local information needs and global information needs. Users interested in a global broad-brush approach will probably need a very high-level vocabulary and very generalized searching, whereas users interested in data at the local level would really want to drill down and get the fine-tuned details.

Using an example of one user group — academic researchers — the working group determined that the kinds of data with which this user group would be involved could include remote sensing images, GIS data sets, etc. Their information needs would include high-quality metadata so that they knew what data they were accessing and whether they could trust it for their purposes.

Users needed bandwidth for downloading very large data sets and perhaps even hardware and software for special clients or special machines to process the information. They might need to be trained in the use of tools and methodologies. They need to have a smooth path toward authentication. Any other access constraints need to be made obvious. Users might possibly need other kinds of tools like gazetteers and other controlled vocabulary.

Typical information-seeking behaviors are to browse or to search geospatially and temporally, as well as taxonomically. Users tend to browse online with the intent of downloading and then using and manipulating the downloaded information offline. However, we recognize that very soon the online computing environment will evolve to a point where manipulation, visualization and analysis could be done on the Internet using Internet map servers and similar devices.

Large groups of the general public — students and teachers at all levels, recreational users of various kinds, birders, gardeners, ecotourists, hunters and fishers, as well as the companies that support these people — are among users of biodiversity information, reported **Caroline Eastman** of the University of South Carolina. The list of user groups expands to includes conservation organizations, funding agencies, wildlife and land managers, and industries — site location, land development, mitigation and restoration, etc. A lot of these folks are willing to pay good money for at least some of this information. (In some cases, however, even if information is available, some of the potential users — e.g., local policy-makers responsible for land development — do not necessarily use it.)

We need to be worrying not just about today's users, but tomorrow's users. We need a call to action to preserve the historical data for which we are responsible today.

What do users want? At least some communities want a real-time global biodiversity monitoring system (at a small scale). Most of the communities will be satisfied with at least a subset of such a system.

The challenges to accomplishing such a system are:

- vocabulary problems
- lack of standards
- inadequate indexing and abstracting
- gaps in data
- errors/variations in data quality
- hard to locate information
- hard to use information/poor interfaces
- privacy/security issues
- problems with interoperability
- data migration challenges
- system changes

Robert Colwell, University of Connecticut, presented his working group's ideas of putting together different kinds of users that normally may not be thought of together. He made the point that data providers are also data users. If MetaDiversity II is a meeting about users, it also has to be about providers. Indigenous people are an example of both users and providers.

There are some categories of planners that hope no biodiversity data is available for a particular area; they want to build a freeway or parking lot on a piece of land or sell it to the timber industry by the millions of acres for clear cutting. And there are planners that hope lots of data are there. These are people trying to plan national parks or protected areas or to get together the data to protect the place that is being threatened by the other kind of planners. The industry of consultants that surrounds both those uses is another heavy user of biodiversity data.

Some of the information that is needed starts with remote sensing data; that is the biggest scale data we have. But remote sensing data is not very useful unless you know what it is that is being sensed down on the ground. Then there is the problem of looking at things on a finer scale to examine at the results of the sensing and to validate their predictions.

At the local level, a fisherman might want to know where to go to find large-mouth bass. An answer to that question could be found from the very same data sets we use for conservation planning or for other kinds of larger scale use. The public, in general, including amateurs, educators, sports people (recreation users), has no way of knowing whether the data is validated and confirmed or worthless. They don't know the peer-review system or one journal from another, but they also need good information. Therefore, one of the ways that our professional community can help is giving credibility or denying credibility to data.

Taxonomic gaps and geographic gaps in data are extreme. There are parts of the world where virtually nothing is known about most groups. To understand the world's biodiversity, there is still just a huge amount of information gathering that needs to be done.

CONTINUING DELIBERATIONS

On the second day of MetaDiversity II, the working groups were charged with focusing on specific needs — systems, content, and functionality — of the active research community and the needs of funding agencies that support their research. The working groups hypothesized about the ideal research environment that will enable researchers and policy-makers to accomplish their tasks. At day's end, each group came forth with recommendations for the National Biological Information Infrastructure in building this ideal type of knowledge environment.

Maureen Kelly noted, "This is an amorphous problem and the needs are very diverse." Kelly offered her distillation, as an outsider, of some of the issues. Biodiversity represents a potentially powerful intersection of diverse communities of interest. This is both good news and bad news. The bad news is: the more you try to deal with the data details, the more the differences surface and start to pull the communities apart. The good news is: there are identifiable points of intersection among the communities of interest. By focusing on these points of intersection, it is possible to achieve a functional level of interoperability for the various resources.

Based on working group discussions, Kelly suggested that some of the high-level points of intersection are issues that relate to geospatial information; time/chronology of collections and data sources; and "the entity" — the name, the taxon, the specimen that is being collected — and information that pertains to it. Other requirements include:

- Additional information associated with identifying the source of the information and the purposes, with the kind of detail that can be used for allowing the user to credential the information;
- Information about the nature of the content: the fields, the formats, and the visual and bibliographic information that supports it;
- Information about the rights to use that information, and restrictions on use, and giving recognition for use.

Working group participants discussed the need to gain a better understanding of the communities themselves that are working with these collections. The diverse needs of these communities need to be recognized and respected. This suggests building an inventory of use cases, the kinds of applications using this data. Thought also needs to be given to the downstream use.

Gary Waggoner, U.S. Geological Survey, said his working group agreed it was important to build a community early on between biologists, biodiversity biologists, computer scientists, information specialists, and social scientists to help establish a broad community endorsement and a common vision and to begin to get members of these various communities working together.

The group recommended expanding access to natural collections globally in a reliable and comprehensive manner. It is very important to develop an advocacy group or an advocacy attitude within the community to promote the use of biodiversity data in education and resources management, conservation activities, land use management, and research, as well as with funding entities (government state and local, nongovernmental organizations, international organizations, and the commercial/corporate world).

Working group participants also discussed the need for teaching curricula for biologists, computer scientists, information specialists, and even social scientists, similar to the way GIS as a tool has been incorporated into the training of a host of different people from different disciplines. There is a need for cross-training and encouraging the development of biological informatics curricula and training.

Scientists need to work with potential funding entities to communicate their needs and priorities, and to discuss how information should be made available when funds are awarded.

It would be a worthwhile activity to sponsor a symposium with ... funding organizations to educate them about some of the global biodiversity issues, the status of existing biodiversity resources, the access problems, and the opportunities for funding projects aimed at solving some of these large biodiversity issues.

Scientists also need to understand the basis for competing for support from these various funding agencies. It would be a worthwhile activity to sponsor a symposium with foundations and interested other potential funding organizations to educate them about some of the global biodiversity issues, the status of existing biodiversity resources, the access problems, and the opportunities for funding projects aimed at solving some of these large biodiversity issues.

In her general comments, **Caroline Eastman,** University of South Carolina, noted that the discussions in her group centered on interests and concerns that would apply across the board to all research communities in science and, in many cases, other areas. Access to literature at a reasonable cost (if not free) was one example.

What is distinctive about biodiversity is that it is specimen-based, to a large extent, while it is also field-based. There is much interest in historical data, and there is a need for global information. A lot of species do not stay in the same place; they move around. Also, the area has a wide variety of public reactions — which can range from indifference to involvement — and interests. Many of the issues are controversial, e.g., endangered species. Public officials may need better information; on the other hand, they may have all the information they want but not care about good science.

Rob Colwell, University of Connecticut, recounted his group's discussion of the urgent need to digitize legacy paper data and specimen data — photographic images, slides in particular, and prints from individual scientists. Protocols and tools to deal efficiently with digitizing those images would be welcome.

Colwell's group called for a workshop among key biodiversity producers, online providers, and aggregators to develop consensus strategies, protocols, and standards for information/data exchange or aggregation in real time, online, from distributed sources (i.e., not warehoused).

COMMON THREADS

Workshop Recommendations

A number of common themes resonated through the discussions of the various working groups. As a group, then, the MetaDiversity II participants offer the following as recommendations for the further development of the National Biological Information Infrastructure in particular, and for the MetaDiversity community in general.

- Inform and educate collectors of data. Articulate the principles of open sources and open access. Give them guidelines and tools for collecting, managing, and archiving data and provide incentives for them to do so.
- Encourage both scientists and members of the civil society — amateurs, ecotourists, birders, gardeners, fishermen, hunters, and life-long learners — to participate in online data collections and to support funding of such efforts. Develop and make available training materials and tools for parataxonomists.
- Involve the children. Prepare the biodiversity researchers of the future.
- Raise the consciousness of the public on the value of online biodiversity resources.
- Develop a more coordinated approach to funding and examine how we can ensure long-term funding streams. Build better bridges to foundations and funding agencies.
- Improve access to biodiversity resources in developing countries and around the world.
- Provide wider access to natural history collections all over the world in a reliable and comprehensive manner. Mobilize sources to digitize and maintain collections, including more support for saving and relocating orphan collections.
- Develop knowledge navigation tools and resources, including authority sources, thesauri, digital gazetteers, rules about data relationships, data mining and analysis software, and systems for attribution, accreditation, and rights management. Develop better visualization tools, virtual laboratories, and ways of dealing with three-dimensional objects, such as specimens, over the web.
- Provide access to a reliable, credible, concept-based, universal taxonomy, a “Catalog of Known Species.” To support that, create a system of unique identifiers that computer systems can use to link information on the same species.
- Maintain web-accessible directories or inventories of sources of information

about software tools, services (particularly taxonomic services), training, biodiversity and biological informatics experts, biodiversity literature, and funding sources.

- Establish data archives to preserve orphan data and to ensure the long-term availability of data, including unpublished data.
- Provide a facility to maintain content in accessible, contemporary technologies, migrating or recapturing legacy data into current technologies as needed.
- Endorse XML and other languages, protocols, etc., that support interoperability across platforms and communication among distributed systems. Develop conversion tools for data to support integration.
- Support, participate in, and build on such global biodiversity systems as the GBIF, the Natural Science Collections Alliance, the National Geospatial Data Infrastructure, and the Global Geospatial Data Infrastructure, among others.
- Develop new technologies, including PDAs that have biodiversity data entry ability, online field guides, smart cameras that tag pictures automatically, better remote sensing tools, etc.
- Convene a workshop at which key biodiversity producers, online providers, and aggregators can develop consensus strategies, protocols, or standards for information/data exchange or aggregation in real-time, online, from distributed sources (i.e., not warehoused).
- Develop publication mechanisms.
- Convene MetaDiversity III to continue the development of this particular community.

These and other pressing issues must be addressed in order to improve access to biodiversity information globally, a necessary prerequisite for preserving our planet's natural heritage through an informed citizenry. By offering this set of recommendations, the participants of MetaDiversity II continue the conversation concerning how to bring the "best science" to all the constituent parts of the biodiversity community. This particular part of the community will continue the discussion … at MetaDiversity III.

APPENDIX A: PROGRAM

MetaDiversity II:
Assessing the Information Requirements of the Biodiversity Community

June 25-26, 2001
Charleston, South Carolina

Sunday, June 24, 2001

4:00 p.m. – 6:00 p.m.	Welcoming Reception

Monday, June 25, 2001

8:15 am – 9:00	Continental Breakfast	
9:00 – 9:05	**Welcome** Brian Sweet, Past President, NFAIS and Barbara Bauldock, USGS	
9:05 – 9:15	**Recommendations from MetaDiversity I and Objective of MetaDiversity II** Jill O'Neill, NFAIS	
9:15 – 10:15	**What Do We Know About Scientists' Use of Information in General? The Role of the Literature and Other Information Sources** Carol Tenopir, University of Tennessee–Knoxville Don King, King Research Tenopir and King are authors of *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*	
10:15 – 10:30	Break	
10:30 – 11:00	**Response from Peer Review Panel**	
11:00 – 12:30 pm	**Presentations on:**	**CONABIO**, Jorge Soberón **DiscoverLife.Org**, Jim Pickering **ITIS**, Janet Gomon
12:30 – 1:45	Buffet Luncheon	

1:45 – 2:30	**What Do We Know About Scientists' Use of Numeric and Other Kinds of Data?** Paul Uhlir, National Academy of Science	
2:30 – 2:45	Break	
2:45 – 4:30	Break-out groups discuss how these general findings relate to biodiversity information systems and prepare a brief report on the three key topics of this symposium. • Identify places where the needs of users in field of biodiversity are being met and how they are being met; • Identify gaps in biodiversity information needed to support users; • Identify population of priority potential users of biodiversity information.	
4:30 – 5:15	Report-out from the Discussion Groups	
5:30 – 6:30	Reception	
6:30 – 8:30	**Seated Dinner with Keynote Address** Meredith Lane, Academy of Natural Sciences	

Tuesday, June 26, 2001 Hibernian Hall

8:00 am – 8:30	Continental Breakfast	
8:30 – 10:00	**Presentations on:**	**Biota**, Robert Colwell **ABI**, Mary Klein **ETI**, Peter Schalk **LTER**, James Brunt
10:00 – 10:30	Break	
10:30 – 11:00	**Response from Peer Review Panel**	
11:00 – 11:45	Discussion Groups begin development of recommendations	
11:45 – 1:00 pm	Buffet Luncheon	
1:00 – 2:30	**Presentations on:**	**Academic Press**, Ken Metzner **JSTOR**, Sherry Aschenbrenner **BioOne**, Heather Joseph
2:30 – 3:00	Break	
3:00 – 3:30	**Response from Peer Review Panel**	
3:30 – 4:30	Discussion Groups complete development of recommendations	
4:30 – 5:00	Report-out and Wrap-up	

APPENDIX B: PARTICIPANTS

Sharolyn Aschenbrenner
Director of User Services
JSTOR
emjocor@umich.edu

Barbara Bauldock
Director, International Biological Informatics Program
USGS/BRD
barbara_bauldock@usgs.gov

James W. Brunt
Associate Director for Information Management
Long Term Ecological Research Network (LTER)
jbrunt@lternet.edu

Bonnie C. Carroll
Consultant to USGS/BRD
Information International Associates, Inc.
bcarroll@infointl.com

C. Ronald Carroll
Director, Institute of Ecology
University of Georgia
rcarroll@arches.uga.edu

Gail Clement
Information Technology Coordinator
USGS Florida Caribbean Science Center
gail_clement@usgs.gov

Robert K. Colwell
Board of Trustees
Distinguished Professor
University of Connecticut
colwell@ucann.edu

Gladys Cotter
Associate Chief Biologist for Information
USGS/BRD
gladys_cotter@usgs.gov

Caroline M. Eastman
Professor
University of South Carolina
eastman@usgs.gov

Thomas C. Edwards, Jr.
Research Ecologist
USGS/BRD, Utah Cooperative Research Unit
tce@nr.usu.edu

Alvaro Espinel
Program Manager, Information Tools
Conservation International, Center for Applied Biodiversity Science
a.espinel@conservation.org

Jacob Faibisch
International Assn. of Fisheries and Wildlife
jacobf@sso.org

Janet R. Gomon
Deputy Director, Integrated Taxonomic Information System
Smithsonian Institution
Gomon.janet@nmnh.si.edu

Andrea Grosse
Biodiversity Information Specialist
USGS/BRD
agrosse@usgs.gov

Joel K. Hammond
Director, Product Database Development
BIOSIS
jkhammond@mail.biosis.org

Preston Hardison
Director
Inter-Governmental Relations for Natural Resources
prestonh@home.com

Eva M. Hedrick
Project Manager, Database Quality Engineering
Chemical Abstracts Service
ehedrick@cas.org

Walter Jetz
University of Oxford, Zoology Department
Walter.jetz@zoo.ox.ac.uk

Heather Joseph
President
BioOne
heather@arl.org

Karen M. Kaye
Information Systems Coordinator
USGS/NBII
karen_kaye@usgs.gov

Maureen C. Kelly
Consultant
mckelly@ix.netcom.com

Donald W. King
Research Professor
University of Pittsburgh
dwking@mail.sis.pitt.edu

Mary Klein
VP Natural Heritage Network
Association for Biodiversity Information
mklein@abi.org

Eric Landis
Natural Resources Information Management
elandis@ix.netcom.com

Meredith A. Lane
Sr. VP for Science and VP for the Biodiversity Research Group
Academy of Natural Sciences
lane@ansp.org

Mark Leiby
Research Scientist
FWC-Florida Marine Research Institute
Mark.leiby@fwc.state.fl.us

Tony Llewellyn
Managing Director
CABI Publishing
t.llewellyn@cabi.org

Robert Magill
Director of Research
Missouri Botanical Garden
Bob.magill@mobot.org

Amy Martin
Online Service Coordinator
Virginia Department of Game & Inland Fisheries
amartin@dgif.state.va.us

Roger McManus
Director of Conservation and Sustainable Development
Missouri Botanical Garden
Roger.mcmanus@mobot.org

A. W. Kenneth Metzner
Vice President, Electronic Publishing Platforms
Worldwide STM Group
Harcourt, Inc.
kmetzner@harcourt.com

Thomas D. Moritz
Director, Library Services
American Museum of Natural History
tmoritz@amnh.org

M. P. Mulligan
USGS Center for Biological Informatics
mike_mulligan@usgs.gov

Jill O'Neill
Director of Planning and Communications
NFAIS
jilloneill@nfais.org

Nick C. Parker
Unit Leader
USGS/BRD, Texas Cooperative Fish & Wildlife Research Unit
nparker@ttu.edu

John Pickering
University of Georgia
pick@discoverlife.org

Dwayne E. Porter
Assistant Professor and Director, GIP Laboratory
Baruch Institute and Norman J. Arnold School of Public Health University of South Carolina
porter@sc.edu

John H. Porter
VCR/LTER
University of Virginia
jporter@lternet.edu

Roberta Y. Rand
Director of the Library
Rosenstiel School of Marine and Atmospheric Science
rrand@rsmas.miami.edu

Diane Riggs
Project Coordinator
USGS/BRD
Driggs01@fiu.edu

Peter H. Schalk
Director, ETI Biodiversity Center
University of Amsterdam
peter@eti.uva.nl

Ralph L. Scott
Director, Project and Program Development
U.S. Department of Energy, Office of Scientific & Technical Information
scottrl@osti.gov

Annie Simpson
Technical Information Specialist
USGS/NBII
asimpson@usgs.gov

Jorge Soberón
Executive Secretary
CONABIO
jsoberon@xolo.conabio.gob.mx

Susan Stitt
Biologist
USGS Center for Biological Informatics
susan_stitt@usgs.gov

Pete Suthard
Supervising Technical Information Specialist
Defense Technical Information Center
psuthard@dtic.mil

Brian Sweet
Director, Publisher Relations
Ovid Technologies, Inc.
brian_sweet@ovid.com

Carol Tenopir
Professor
University of Tennessee
ctenopir@utk.edu

Robert Turner
Tech. Director, NBII Southern Appalachian Information Node
Oak Ridge National Laboratory/University of Tennessee/SAMAB
rtt@ornl.gov or
rsturner@utk.edu

Paul F. Uhlir
Director, International S&T Information Programs
The National Academies
puhlir@nas.edu

Gary S. Waggoner
NBII Biodiversity Coordinator
USGS
gary_waggoner@usgs.gov

Jeff Waldon
Assistant Director
Conservation Management Institute
fwiexchg@vt.edu or
sheliar@vt.edu

Samuel Walker
USC-The Baruch Institute
sam@inlet.geol.sc.edu

Daniel U. Wilde
NERAC, Inc.
duwilde@mail.nerac.com

Crispen Wilson
BCIS Programme Manager
Conservation International
c.wilson@conservation.org

APPENDIX C: WORKSHEETS

User Group Analysis

Target Group EXAMPLE: Botanists	**Description of Activities** (Types of data collected, where and how is data captured, what is captured)	**Information Needs** (Types of resources, data this user group would want to be able to access)	**Information-seeking Behaviors** (How does this group think about their information? What are the constraints faced?)

Hypothesize about the knowledge environment that would enable researchers to accomplish the tasks that they have. Information types and formats that support the needs and behaviors:

Need or Behavior	**Supported By** (Browsing or visualization or search engine)	**Information Type or Format**