by Daniel Edelstein

BioMed/MS Beagle* **2/12/2000 edition**

(* = Elseveir Science Publishing's subscriber Web site for science professionals)

Giant “Yellow Pages” of Life on Earth Goes Online

Abstract

Later this year, a database that pools the earth’s animals, plants and microorganisms will be accessible via the Internet. As the first-ever source to provide scientists data relating to all the world’s cataloged species, the electronic network will be called the Global Biodiversity Information Facility. Instant access to natural history information is expected to help scientists track and link species while also allowing researchers to locate, retrieve and synthesize data from multiple sources as they access climatic, phylogenetic, evolutionary, and molecular data to help create conservation management plans.

*** * * * * ***

Imagine a fisheries resource manager collecting samples of fish in various streams throughout a watershed area. Upon returning to the office, he accesses an online fish-identification program. The manager quickly determines his catch includes several alien species. Another Internet database tells him the native ranges of these species, their reproductive requirements, ecological preferences, natural predators and parasites. Now the manager deduces only one of the alien species is likely to survive and spread throughout the watershed.

Concurrently, he combines environmental data specific to each fish’s native range that has been measured by satellite passes for the past 20 years with the habitat conditions of the other areas into which they have been introduced — and thereby

interprets the data to ascertain which management strategy is most likely to succeed in the current scenario. In a mere few hours, the resource manager has used a rich network of online databases to control the populations of invading fish species before they may grow out of control.

Does this kind of online database management tool sound farfetched and impossible? It's not. In fact, scientists worldwide will soon be able to use this kind of electronic network "yellow pages" that highlights all of the world's plants, animals and microorganisms. Called the *Global Biodiversity Information Facility* (GBIF), the linked databases will be the first to include a definitive list of the world's 1.5 million named species. This initiative, sponsored by the Working Group on Biological Informatics of the Megascience Forum of the Organization for Economic Cooperation and Development (OECD), is the largest of several that purports to make collections-based research and information accessible to both general and scientific communities by digitizing it and storing it on networked databases.

A Global Resource

"The Global Biodiversity Information Facility will be a fantastic resource for everybody," said Dr. James Edwards, an assistant director at the National Science Foundation in Washington and one of several scientists involved in the project among the OECD's 29 industrialized nations.

For people having trouble understanding how to envision the massive scope of the GBIF, Steve Young, a computer specialist working on the project at the Environmental Protection Agency suggests they, "Bear in mind that the GBIF is not a monolithic facility like a new giant telescope or particle accelerator. It will be an organic, distributed, growing, evolving
facility, much like the World Wide Web itself."

GBIF's Arrival: Step-by-Step

Edwards said the GBIF will become functional in stages. In March, a Web site will open as part of the proposal's goal of creating a home page for every species on the planet. Links will then be created to the existing data and integrate information from other online projects that maintain inventories of the world's species. Edwards expects some of these shared database links to be operational by January, 2001, though more in-depth integrated searches may not be possible until 2002 and beyond.

The databases that operate jointly within the GBIF will remain the property of the institutions that own and provide the data and link them to the Internet. "This seemed more appropriate and more practical than starting a single central database for biological information from scratch," said Michael Oborne, deputy director for the OECD in Paris.

Thomas Lovejoy, chief biodiversity adviser to the World Bank and one of the planners of the GBIF project, suggests its greatest appeal lies in updating and validating the worldwide efforts of natural history scientists: "Natural history began hundreds of years ago and has stayed behind" the progress shown in many other younger core science areas (such as molecular biology and genome research) that have always used network databases to share their research.

Rapid Transit

Faced with an immediate need for implementing conservation plans in an effort to save dwindling populations of flora and fauna worldwide, scientists will be able to use the GBIF to rapidly obtain information. Scientists, for example, will simplify their searches and conclusions by identifying data sets for a species or family with the click of a mouse.

"Right now, if I wanted to know all the specimens of a particular bird," Edwards said, "I would have to write to all the museums that may have a specimen, then go to the museums myself and look at labels and cards. Our initial focus is to get the data

from the tags and the drawers into the computer and, where possible, include images."

And the Poor Get Richer

More than providing benefits to scientists, professors, researchers and biology students, the massive and comprehensive GBIF system promises to assist poorer countries that are rich in diversity of species but have no species cataloging system of their own. These developing countries currently host most of the world's living plant and animal species, Edwards said, but the collections of them are often housed in databases and natural history collections found in northern Europe and North American universities, zoos and museums. Strange as it may sound, then, a field ornithologist from South America seeking information about the native tanagers he is observing daily may have to travel overseas to access an archival collection housed in a northern hemisphere locale. For this reason, Lovejoy emphasizes, "the idea (of creating GBIF) is to pull all of this together and make it available to everybody" as a means of digitizing the data and supplying them back for utilization in the countries of origin.

Challenging Times

Initially, Young said the GBIF is expecting to have a staff of around 10 employees and open with the developed countries paying more to join than developing countries. More complex challenges are expected soon after all involved parties begin sharing their biodiversity databases with the world.

One of the most formidable roadblocks is not unlike the recent cataloging and search engine maladies that plagued the World Wide Web when the general population first began browsing for information. Dozens of search engine companies — among them Yahoo, Alta Vista, Lycos and Excite — were founded and stressed to perform favorably

despite troubling technical questions that arose when independent databases began to be linked.

Edwards does not tremble with fear at the similar complex technical barrier facing the GBIF in its quest to link, organize and track the world's huge volumes of biological data. "Until a few years ago, the technical hurdles in linking disparate datasets were formidable," he said. "In recent years, Web-based search engines have begun to proliferate, and it is now is feasible to do distributed searches across a wide array of different kinds of data and database formats."

Common Bonds and Links

Crucial to overcoming the challenge of linking databases, Dr. Edwards explained, is identifying a common element when a scientist begins a search within GBIF. "For most biodiversity data, a valid species name will provide that common element; this is why GBIF has placed such an emphasis on completing a catalogue of the names of known organisms. For other kinds of data, other common elements must be used; for example, biogeographic data use place names of latitude-longitude coordinates.

"Part of what we'll have to do is to come up not only with the named species but also all their synonyms," Dr. Edwards said. "Now that we try to computerize things, we'll have to weed out the ambiguity that exists. This will be of value not only to scientists but also to policy makers. If people prepare a new law, say to protect a species, there can be no ambiguity."

Shared Knowledge

Is there opposition to the GBIF? — especially from developing countries that may both be wary of foreign intrusion upon their ample supply of natural resources and mindful of past invasions by multinational companies looking to obtain virgin resources from the vast tropical gene pool. Not yet, say the project's developers, with Young reminding:

“One of the key aspirations of many of the GBIF organizers is to support the repatriation of information back to source countries; in other words, to share back with the countries of origin of biological specimens the information about those specimens, so that the knowledge is shared. We believe this will be well received by the developing countries. If they are able to access the knowledge, without incurring all the costs of curating large collections over a vast time, then this should be very beneficial to them.”

The projects organizers are also are quick to point out that great pains have been made to establish constructive relationships with existing governmental and non-governmental activities that currently work in the informatics and conservation field. To date, these strategic and diplomacy efforts have included inviting the involvement of the Convention on Biodiversity and its subsidiary bodies, the Clearinghouse Mechanisms and the Subsidiary Body on Scientific, Technical, and Technological Advice. The Convention will have a free non-voting seat on GBIF's Governing Board.

Indeed, scientists worldwide are expected to use the GBIF regularly as they “find and use biodiversity data from other scientists and organizations and use the GBIF as a way of ‘publishing’ and exchanging their own scientific findings with others around the world,” according to Glady Cotter, a USGS associate chief biologist of information.

Of greater concern to Cotter may be the incompatibility of GBIF systems with existing biodiversity and ecosystem databases. In citing her work for the United States current version of a biodiversity database, the National Biological Information Infrastructure (NBII), she emphasizes the key to successful exchange of GBIF data among scientists will be maintaining consistent infrastructure standards when listing regional, national, and global biological information. These catalogs currently exist, for example, as part of online Natural Heritage Programs that most states already operate, and in other regional biodiversity listings.

The NBII, in fact, will be the first building block of the GBIF database, along with similar worldwide projects. Some of these online programs, including current ones in Australia, Britain, and Canada, offer biological information on their main domestic species, but these countries' biodiversity inventories are also not yet complete. In addition, many of the world's historical plant and animal collections, such as those in Britain, France, Germany and Russia, often date back to the 18th century, and have not been entered into computer records.

"Creating a GBIF is connecting those (worldwide) pieces into a larger whole; it's not preempting (them) in the least," Lovejoy suggests.

Integrated Virtues and Extended Features

Other GBIF developers are also unanimous in their enthusiasm for extended features that will eventually allow the integrated biological network to be popular among professionals in fields apart from the environment, including economic planners and industries such as pharmaceutical companies that do biological prospecting. These features, called "vertical links," will assist management decisions based on linking the species databases with others dealing with soil, climate, human populations, animal distribution, vegetational patterns, watersheds and others.

A simplified "vertical link," Cotter mentions, might challenge scientists to use the GBIF to locate, retrieve and synthesize data from multiple sources as they investigate the effect of a pesticide on an amphibian known to be declining in population size in certain areas of North America. Using the GBIF network, the results of the study could then help resource managers identify the climates where the species lives, its competitors in an ecosystem and how all of them will be affected by the pesticide.

Edwards declares his own growing optimism for the value of the GBIF is partially based on a current successful real-world integration of environmental information in Sweden. There, he said, "Use of environmental information is contributing to the development of environmental policies, assessment of environmental impact, and developments

leading to sustainability. For example, information about plant and animal species in Sweden has been integrated with climate models to predict the distribution of a range of species under a number of climate change scenarios developed from Global Climate Models. This information has then been integrated with other information about vegetation for use in the development of species management plans.”