way to creating the "invisible magic" needed to locate and track data for precisely the right organism in desktop biodiversity systems on the Internet.

Robots and knowledge integration. The MultiFlora system (*30*) being developed at the University of Manchester and the Natural History Museum, London, uses Information Extraction (IE) to seek information on the same species from parallel unstructured text resources. Information from the various sources discovered is then returned, using XML, and assembled into a single database, with some interesting features for handling variable and conflicting data. The idea is that redundancies between sources may allow the system to create accurate databases despite some of the shortcomings of IE techniques.

A further biodiversity analytical system is in development at the Natural History Museum, London. The WORLDMAP system (*31*) can be used with distribution data sets to plot measured species biodiversity distribution patterns and to highlight hotspots and areas of endemicity. Of interest is the array of biodiversity measures provided, including the much debated taxic measures that incorporate distances over the phylogeny.

Startling as all these developments may be, they might just be the tip of an iceberg, preceding undreamt-of models in the coming century. Certainly there are those who expect the Internet, as seen by biologists, to become one giant global biodiversity information system. Even biologists who spend a lifetime of travel and fieldwork cannot observe the whole. But as an abstraction, could global biodiversity come to exist, modeled and visualized, on the Internet as nowhere else?

References and Notes

1. Species 2000 (www.sp2000.org).
2. J. L. Edwards, M. A. Lane, E. S. Nielsen, *Science* **289**, 2312 (2000).
3. Environment Australia Online (www.environment.gov.au/search/search.html).
4. BioCISE project (www.bgbm.fu-berlin.de/biocise/default.htm#).
5. The Species Analyst (http://habanero.nhm.ukans.edu).
6. Darwin Core metadata standard (http://habanero.nhm.ukans.edu/Z.X).
7. R. Leow and K. Taylor, in *Proceedings of the 12th International Conference on Scientific and Statistical Database Management*, O. Gunther, Ed. (IEEE Computer Society Press, Los Alamitos, CA, 2000), pp. 25–38.
8. ENHSIN (www.nhm.ac.uk/science/rco/enhsin).
9. ITIS (www.itis.usda.gov).
10. URMO (www2.eti.uva.nl/database/urmo/default.html).
11. IOPI Global Plant Checklist (http://bgbm3.bgbm.fu-berlin.de/iopi/gpc).
12. F. A. Bisby, in *Designs for a Global Plant Species Information System*, F. A. Bisby, G. F. Russell, R. J. Pankhurst, Eds. (Oxford Univ. Press, Oxford, 1993), pp. 145–157.
13. ILDIS LegumeWeb (www.ildis.org).
14. FishBase (www.fishbase.org).
15. BIOS (www-sp2000ao.nies.go.jp/bios/index.html).
16. Species 2000 Interoperability Coordination Environment (SPICE for Species 2000) project (www.systematics.reading.ac.uk/spice).
17. GBIF (www.gbif.org).
18. SINGER (www.singer.cgiar.org).
19. FloraBase (http://florabase.calm.wa.gov.au).
20. ERMS (http://erms.biol.soton.ac.uk).
21. CephBase (www.cephbase.dal.ca).
22. ALICE Software (http://dialspace.dial.pipex.com/town/square/fd95).
23. J. H. Beach, S. Pramanik, J. H. Beaman, in *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*, R. Fortuner, Ed. (Johns Hopkins Univ. Press, Baltimore, 1993), pp. 241–252.
24. Y. Zhong, S. Jung, S. Pramanik, J. H. Beaman, *Taxon* **45**, 223 (1996).
25. W. G. Berendsohn, *Taxon* **44**, 207 (1995).
26. M. Pullan, M. Watson, J. Kennedy, C. Reguenaud, R. Hyam, *Taxon* **49**, 55 (2000).
27. G. P. Yakovlev, Y. R. Roskov, A. K. Sytin, S. A. Jezniakowsky, *Legumes of Northern Eurasia* [CD-ROM] (ILDIS Northern Eurasia, St. Petersburg, 1998).
28. LITCHI Project (http://litchi.biol.soton.ac.uk).
29. A. C. Jones *et al.*, in *Proceedings of the 12th International Conference on Scientific and Statistical Database Management*, O. Gunther, Ed. (IEEE Computer Society Press, Los Alamitos, CA, 2000), pp. 3–13.
30. MultiFlora (www.cs.man.ac.uk/ai/MultiFlora).
31. WORLDMAP (www.nhm.ac.uk/science/projects/worldmap).
32. F. A. Bisby and P. M. Smith, Eds., *Species 2000 Project Plan* (Species 2000 and UNEP, Southampton, UK, 1996).

33. I thank J. Heald and S. Brandt for comments on the text and R. White, J. Beach, M. Scoble, and J. Croft for assistance with particular items.

VIEWPOINT

Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop

James L. Edwards,[1]* Meredith A. Lane,[2] Ebbe S. Nielsen[3]

Data about biodiversity are either scattered in many databases or reside on paper or other media not amenable to interactive searching. The Global Biodiversity Information Facility (GBIF) is a framework for facilitating the digitization of biodiversity data and for making interoperable an as-yet-unknown number of biodiversity databases that are distributed around the globe. In concert with other existing efforts, GBIF will catalyze the completion of a Catalog of the Names of Known Organisms and will develop search engines to mine the vast quantities of biodiversity data. It will be an outstanding tool for scientists, natural resource managers, and policy-makers.

Biodiversity is distributed all over the Earth, with the highest concentration in tropical regions, especially in developing countries, and in the oceans. In contrast, scientific information about biodiversity is largely concentrated in major centers in developed countries, especially in the scientific collections of the world's natural history museums, herbaria, and microorganismal repositories. At present, it is more likely that information on the plants of a particular part of Africa is stored in an herbarium in Europe, for example, than in its source country. Approximately 3 billion specimens of organisms of all types are held in the natural history collections of the world (*1*). Each of these specimens has associated data, including, at the minimum, the scientific name of the specimen, when and where it was collected, and by whom. Many specimens also have other kinds of associated information, including pointers to other physical samples derived from the specimen (e.g., frozen tissues, DNA extracts, hosts, parasites), photographs, recordings of mating calls or other behavior, the field notes of the collector(s), and a wide range of other data.

In toto, then, there is an enormous amount of information already collected about the world's biodiversity. However, to date most of this information has not been digitized. Thus, in most cases, the only way a potential user can find out about the data is to travel physically to the place where the specimen is housed or to contact the repository where a relevant specimen may be housed and ask to borrow it (and its associated data).

The sustainable use and management of biodiversity will require that information about it be available when and where that information is needed by decision-makers and scientists alike. Because biodiversity information is not immediately at hand, it is often not applied in policy or management decisions that affect the organisms involved, nor is that information readily accessible by

[1]Directorate for Biological Sciences, National Science Foundation, Arlington, VA 22230, USA. [2]Academy of Natural Sciences, Philadelphia, PA 19103, USA. [3]Australian National Insect Collection, CSIRO Entomology, Canberra ACT 2601, Australia.

*To whom correspondence should be addressed. E-mail: jledward@nsf.gov

research scientists. To remedy this situation, a number of activities have been started at various places around the world to set up biodiversity databases. A quick glance at the Biological and Biodiversity Web Server (*2*) reveals a wide range of searchable databases, almost all of which are incompatible with each other. But it is precisely through linkages among data from different sources and on different topics that the greatest rewards are likely to be realized.

As noted by Maurer *et al.* (*3*), large sophisticated databases do not just happen; they must be carefully planned, with interoperability in mind. However, that does not mean that all the data need to be stored in a single computer. With only a minimum of metadata information about an online database, recent browser technology such as extensible markup language (XML) will allow a user to search several databases. An example is "Species Analyst" (*4*), which, with the use of XML, can currently search approximately 10 million specimen records from natural history collections located at 13 institutions worldwide, pull the information together, and produce secondary analyses such as species distribution maps and predictive modeling, all in real time. A range of other powerful and user-friendly applications for the capture, management, and sharing of biodiversity data are now available, together with tools for mining (extracting relevant data), analysis, and visualization of biodiversity data [see "Links" at (*5*)].

Thus, the situation is ripe for an organized attempt to bring the massive amount of biodiversity data located in natural history collections to the desktop of any user. This is the goal of the Global Biodiversity Information Facility (GBIF) (*5*), a proposed mechanism for coordinating and making interoperable biodiversity databases that are distributed worldwide. Each database will continue to be owned and managed by the institution that holds the specimens which serve as vouchers to verify the database records. GBIF will provide a mechanism for database affiliation, will promote and enhance the development of the standards required for interoperability, and will leverage development of various specialized search engines for accessing the data.

The formation of GBIF is being overseen by an Interim Steering Committee comprising representatives from at least 25 countries (*6*). An invitation to become involved in the establishment and growth of GBIF has recently been sent to the science ministers of all other countries.

At the heart of GBIF will be a catalog of the scientific names of all the world's species (Fig. 1). About 1.8 million species have been given scientific names to date (*7*). (It is estimated that anywhere from 3 million to more than 100 million species exist in the world today (*8*), so biodiversity scientists still have a lot of work to do.) However, many species have been named more than once by different scientists, and, as our knowledge of species has changed over the years, species have often been merged together or split into several new species. Thus, there are many (perhaps three times as many) more scientific names than there are valid species. The scientific community has begun several international projects to sort out the nomenclatural confusion that has resulted; examples are Species 2000 (*9*, *10*), the International Plant Names Index (*11*), and the Integrated Taxonomic Information System (*12*). However, each of these projects has taken a different approach; one of GBIF's primary goals will be to coordinate them and facilitate their rapid completion.

Longer term goals for GBIF are to develop both a digital library of biodiversity knowledge drawn from information available in print libraries and "Species Bank," a compilation of facts about each individual species.

GBIF cannot accomplish this ambitious agenda on its own. Rather, it will join forces with existing activities and programs. At the national level, these include Mexico's Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) (*13*), Costa Rica's Instituto Nacional de Biodiversidad (INBio) (*14*), and the Australian Biodiversity Information Facility (ABIF) (*15*). Close connections are being developed with the Clearing House Mechanism of the Convention on Biological Diversity (*16*) and with DIVERSITAS (*17*), an international program of biodiversity science that is sponsored by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and several of the members of the International Council for Science (ICSU).

In addition, new projects that are explicitly intended to be particular modules within GBIF are being developed. One of these is the Ocean Biogeographic Information System (OBIS) (*18*). Another, in the early stages of development, is a series of linked databases that contain information about the phylogenetic relationships of organisms. The "Tree of Life" project will expand on the current "Tree of Life" Web site (*19*) and will incorporate "TreeBase" (*20*) as well as visualization tools and other components. The information-science research needed to bring this phylogenetic content to the World Wide Web is coming to be called "phyloinformatics."

GBIF will also foster interoperability of biodiversity databases with those in domains outside of species biodiversity, such as sequence and other molecular data, geospatial and climate data, and ecological and ecosystems data. These linkages will enable "data mining" never before possible, and facilitate the exploration of questions that, at present, cannot readily be answered. For instance, "What is the ecological context (location, climate, associated species, etc.) within which a particular gene evolved, where can specimens of the associated species be found, by what other names might those species be known, and are any of them (or taxa phylogenetically related to them) represented by sequences in GenBank?"

The target date for establishing GBIF is early in 2001. GBIF will be open and freely available to anyone with access to the World Wide Web. Most of its activities will be carried out within member countries, supported by their national funding programs. A GBIF secretariat will provide a clearinghouse for information about past, present, and pro-

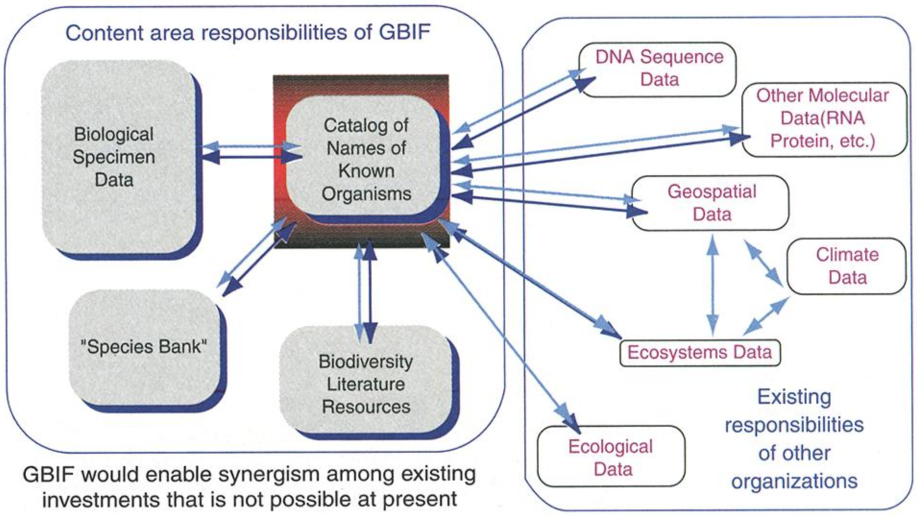

Fig. 1. An electronic Catalog of the Names of Known Organisms will make linkages among many types of biological and nonbiological databases possible. These linkages will allow data mining that cannot be imagined today because, at present, it is difficult if not impossible to uncover correlations among different data sets.

posed projects; promote interoperability among databases by holding standards-setting workshops and developing protocols for linking databases; provide coordination for national funding programs and search for additional funding from international and regional bodies; and facilitate practical applications that demonstrate the utility of biodiversity data for addressing critical social concerns.

GBIF is a grand, interesting experiment in creating a global capacity for access to biodiversity data. Its databases will be distributed widely; at the same time it will encourage cooperation and coherence. It will be global in scale, though implemented nationally and regionally and will be open to participation and benefit by all countries. GBIF will aid in advancing scientific research in a host of areas, including systematics, conservation biology, ecology, agriculture, biomedicine, and environmental management. It will serve the economic and quality-of-life interests of society and will provide a basis from which our knowledge of the natural world can grow rapidly and in a manner that avoids duplication of effort and expenditure.

References and Notes

1. M. A. Lane, *Weaving a Web of Wealth: Biological Informatics for Industry, Science and Health* (Australian Academy of Science, Canberra, 1999).
2. http://www.keil.ukans.edu/
3. S. M. Maurer, R. B. Firestone, C. R. Scriver, *Nature* **405**, 117 (2000).
4. http://habanero.nhm.ukans.edu/
5. www.gbif.org
6. As of September 2000, members of the Interim Steering Committee included Australia, Bangladesh, Belgium, Brazil, Canada, Costa Rica, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Mexico, Netherlands, New Zealand, Poland, Portugal, Slovenia, South Africa, Spain, Sweden, the United Kingdom, and the United States, as well as representatives from the European Commission, the Convention on Biological Diversity, and the Commission for Environmental Cooperation.
7. P. M. Hammond, in *Global Biodiversity Assessment*, V. Heywood, Ed. (Cambridge Univ. Press, Cambridge, 1995), pp. 113–138.
8. R. M. May, in *Nature and Human Society: The Quest for a Sustainable World*, P. H. Raven, Ed. (National Academy Press, Washington, DC, 1999), pp. 30–45.
9. www.sp2000.org/
10. F. A. Bisby, *Science* **289**, 2309 (2000).
11. www.ipni.org/
12. www.itis.usda.gov/
13. www.conabio.gob.mx/
14. www.inbio.ac.cr/
15. www.anbg.gov.au/abrs/abif.htm
16. www.biodiv.org/chm/
17. www.icsu.org/DIVERSITAS
18. D. Malakoff, *Science* **288**, 1575 (2000).
19. http://phylogeny.arizona.edu/tree/phylogeny.html
20. http://herbaria.harvard.edu/treebase/