

Overview of Databases

R. E. Stinner
NSF Center for Integrated Pest Management
North Carolina State University

The preceding abstracts provide a description of many of the databases discussed at the workshop, and additional detail on most of the databases described can be obtained by reviewing the completed data survey information forms on the National Biological Information Infrastructure (NBII) website on invasive species at <http://www.nbii.gov/invasive/workshops/dbsurveys.html>.

Any database on biodiversity should be considered as containing potentially important information for enumerating invasive species and their attributes. The databases reviewed at the workshop represent a small but important fraction of those developed and available. These databases do, however, provide examples of the spectrum of information needed in studying and documenting invasive species.

Regulated and Invasive Species

Among those databases specific to invasive species are several maintained by the USDA's Animal and Plant Health Inspection Service (APHIS): Federal Noxious Weeds, the North American Nonindigenous Arthropod Database, Identified Plant Pests Regulated by APHIS, and the Port Information Network. The latter two databases include only regulated and excluded pests and the information is limited to taxonomic and geographic data. More detailed biological information, together with risk analysis results, can be found in the Exotic Forest Pest Information System for North America being developed by the North American Forestry Commission and available in English, French and Spanish.

Other invasive species databases, such as the Invasive Species of Indian Ocean Islands and the Invasive Species of Pacific Islands, are relatively new efforts and focus on species that threaten biodiversity and conservation values. The World's 100 Worst Invasive Species is a pilot database being developed by the same Global Invasive Species Programme as the previous databases, again with those species threatening to biodiversity as the targets. On a more

local basis, the Connecticut Invasive Plant Task Force is developing a database of brochures, websites, and other literature for the lay public.

Single Species Spread

There are a number of databases that provide detailed information on single species with a history of invasiveness. These databases provide more in-depth information, usually with current status of the species spread. Examples here include the Slow-the-Spread Gypsy Moth database and the Witchweed Management database. Such databases typically also contain historical information allowing analysis of long-term invasions and other distributional data.

Biodiversity/Taxonomy

Almost every museum and taxonomic research unit maintains taxon-specific databases with a focus on biodiversity. Of particular use are survey databases such as the Illinois Natural History Survey Collections and the North American Breeding Bird Survey. One of the most important efforts in this area is IABIN (Inter-American Biodiversity Information Network, see website www.nbii.gov/iabin), a collaborative effort of a number of museums working in biodiversity. This network uses a Z39.50 compliant client and Z servers to allow access to cooperating distributed databases.

Pest/Economic

Because of their economic importance, many taxa have databases detailing their biology, distribution, and life histories. Examples here include the Microbial Germplasm Database, the ROBO Database, and Hymenoptera-Online.

Database Considerations

Many of the above databases do not contain information on invasiveness specifically, but rather data on distributions, biology, and historical movement. The sheer number of databases available force us to

consider issues of database access, integrity, and continuity. Present technology allows us to share information, but compatibility, ownership and database security present formidable challenges.

Compatibility is a complex question incorporating both philosophical problems (naming conventions) and new technologies. The need for vocabulary standards and dictionaries represent only the first steps in dealing with compatibility issues.

New technologies such as XML, thin-clients, and COBRA-compliant software greatly increase our abilities to interact. These technologies are even changing our fundamental definition of databases. The World Wide Web has already done that. The largest databases ever to exist, by far, are the indexed searches of the major Web search engines. They are relatively disorganized and dumb, but XML and the next generations of Web languages and software could well make present day database efforts obsolete.

It is a simple matter today to embed searches (queries) to almost any Web database in another database or Web page. For data on the public Internet, trained experts can access that data and manipulate the resulting query output prior to presentation. At least among government and university workers, we must move to develop protocols to guide decisions on ownership designation and output configuration.

Database Coordination

There needs to be a significant effort to identify what key databases are presently available and which are needed, as well as an analysis of critical information which may be lacking in those databases already available. This need not be an exhaustive search, but rather the identification of those of greatest significance. Many of the databases available are taxonomic in nature. They contain little biological information and do not separate invasive from noninvasive species. However, they are critical in determining species distributions and need to be analyzed through time to present information on invasiveness propensities and pathways.

What is clearly lacking is a blueprint for coordination and use of these myriad databases. The key to pre-

vention and control of invasive species lies in our ability to concurrently access the major invasive species databases. By creating on-demand documents containing all of the current and critical information on a given species, regulatory agencies can prepare management plans for interception, containment, or control within scientifically acceptable time constraints. This same information is also crucial in the development of predictive models for invasive species.

Many of the present databases are housed at land-grant universities on servers with high-speed, Internet access and highly trained information technology experts. The continued and expanded use of such facilities/expertise for database development and maintenance would enable continued access to these databases at modest costs.

All critical databases must be made available on-line. There must be an international effort to develop and maintain on-line database search tools capable of intelligent queries of key databases. Increasing global trade and resultant increasing threat of invasive species require quick and informed actions. Present technology allows this capability, but requires database coordination and cooperation at the highest level to provide funding for both the databases, their integration, and concurrent access to that information.

Such a program should involve not only U.S. government agencies, but should seek partnerships with state and foreign governments, universities, nonprofit organizations, and companies with a stake in preventing the ingress of invasive species. Understanding and managing invasive species requires all stakeholders to commit to the establishment and use of invasive species databases which can be used and interpreted quickly and accurately. We must all agree that *"information is not power; the sharing of information is."*

Models for Database Integration and Use

As we look to the future, we must develop processes that will enable us to share information much more effectively. However, in view of the large number of highly diverse databases and the many different perspectives that the developers of those databases have, it is probably unrealistic to attempt to develop a national plan that would deal with all of the integration, ownership, and use issues. Therefore, efforts to

deal with some of these key issues can best be directed towards efforts to work with small clusters of databases whose developers have close common interests to develop one or more models. One or more such models could then serve as a basis for developing protocols for more extensive integration of databases useful in dealing with invasive species. In view of the large resources being devoted to databases and the relative lack of coordination among them, action

should soon be undertaken to develop models for the integration and use of databases. Databases provide a unique opportunity to link diverse interests and achievements, to enhance understanding, and to build consensus. Therefore, the conduct of some sharply focused facilitated activities involving stakeholders and public agency representatives should be organized as soon as possible.