



[<< Back to Article](#)

Building a Database of Specimens

Stewart Taggart  07.02.01 | 2:00 AM

SYDNEY, Australia -- In a windowless back room of the Australian Museum, there's a place called the "Spirit House."

Thousands of animal specimens are kept there, preserved in jars of alcohol. Most are identified by little more than dusty paper cards or crusty labels --? a tribute to pre-computer-era science.

Around the world, roughly 3 billion such animal and plant specimens sit in places like the Australian Museum,? with no comprehensive electronic means for researchers to share data about them. Now, after nearly five years of negotiations, a major effort is underway to create an international electronic *lingua franca* for this sleeping scientific bounty.

Earlier this month, 18 nations agreed to establish the Global Biodiversity Information Facility (GBIF) in Copenhagen, which will establish a standard method for data exchange about such specimens for researchers in different countries. By developing a kind of online phone book of the world's major plant and animal collections, the potential for new scientific discoveries could be huge.

Among the possibilities?

Bio prospectors could follow up interesting drug leads should compounds in one plant be found to have cousins or distant relations in other lands. More sophisticated conservation programs could be developed as more information on natural environments becomes available across political borders. And plant and animal migration theories could be altered as researchers find new genetic links among animals of different continents.

But the big task right now is just getting the stuff online. As a database project, it could make the Human Genome Project look simple by comparison.

"Man has been collecting biological data for centuries," said Dr. John Curran, assistant chief of the Commonwealth Scientific and Industrial Research Organization's division of Entomology (aka insects). "The genome people had the advantage in that they started collecting and working with their data after computers existed."

To get an idea of the challenges ahead, go back to the Spirit House. There, a 19th-century offline archival system holds sway, based largely on elaborate cursive script and paper tags. Getting it all into data fields parsed by mainframe computers will be a huge task.

And that's just the headache facing the Australian Museum in Sydney, which has an estimated 7 million specimens in its animal collections. In Canberra, the Australian National Insect Collection holds roughly 11 million insect specimens. Australia's seven major herbariums located around the country have nearly 6 million specimens among them.

And this tally is just for a few select institutions in Australia. Similar numbers no doubt exist in other countries with equally large collections. What's more, the databasing task is both quantitative and qualitative.

For instance, among Australia's herbariums, only 40 percent of the data on specimens is currently in any kind of electronic form suitable for inclusion in GBIF, said Dr. Judy West, director of Australia's Center for Plant Biodiversity Research.

The rest of the data still needs to be entered into computers, and much of it lacks a lot of the detail today's researchers want, such as the precise GPS location of where specimens were collected. Collectors in, say, 1850, did not have that technology, said West. Therefore, much of that data may have to be reverse-engineered as closely as possible, she said. Some of Australia's specimens date from as early as 1770.

Nonetheless, the 18 countries that have signed up for the initial stage of the GBIF agree the ramifications of such linked data will be staggering, and the project will require a multidisciplinary? approach once it gets going.

"GBIF will provide an unparalleled source of biodiversity information," said GBIF Chairman Christoph Haeuser, a biodiversity researcher at the Stuttgart State Museum. "People will be able to use this facility from anywhere in the world to access geographical, ecological, genetic and taxonomic information."

The origin of GBIF dates to January 1996, when the Organization for Economic Cooperation and Development (OECD) agreed that more broadly available information on earth's biodiversity was critically important to society, but was being bottled up by database incompatibilities across nations and offline archiving.

GBIF was established in mid-2000, and earlier this month Copenhagen was selected as its headquarters, beating out rival bids from Australia, Spain and Holland. Now that the headquarters location is fixed, the real work can start. And the sooner the better, because many plant and animal collections continue to grow at a rapid rate.

The sooner electronic standards are created, the less work will need to be done down the line to get collections ready.

And as this bread-and-butter work gets done, the real payoff will come as GBIF's databases increasingly are parsed along other databases of similar magnitude, such as those for climate records, environment and geography.

"Once you lash together all this information, who knows what questions we could be asking as scientists?" Curran said. "The other databases are much farther along. Biological data is in many ways the laggard here."