Taxonomic Revival

By putting museum collections online and training students to be computer- and molecular biology–savvy, taxonomists hope their field will thrive in the new millennium

The biodiversity crisis is not just about the perilous state of plants and animals. Accumulated knowledge about each species is also under threat. For several decades, the plight of pandas, whales, woodpeckers, and butterflies has regularly made headlines,

Bioinformatics proponents. (Left to right) Kansas's Jim Beach, Dave Vieglais, Daphne Fautin, Steve Ashe, and Ed Wiley, with some of their collections.

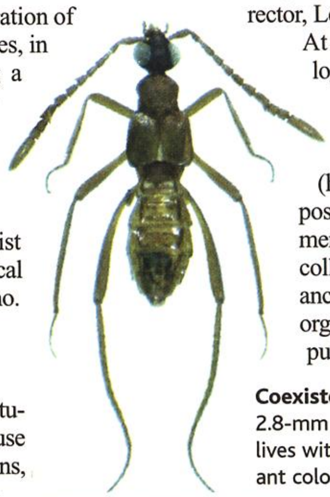
while scores of conservation organizations, government agencies, and private foundations have worked to stem the decline. But only recently has attention turned to protecting the other side of biodiversity.

One "hotspot" of our knowledge of organisms is the drawers and cabinets full of animal hides, bones, bodies, and mounted plant specimens warehoused in natural history museums, herbaria, and what were once the zoology, entomology, and botany departments of universities. Another hotspot is the aging taxonomists and systematists who are retiring, taking with them their in-depth understanding of whole groups of organisms. The preservation of both types of knowledge has finally begun to attract attention from the conservation community, biologists, and—perhaps most important—funding agencies.

Two decades ago, taxonomy and systematics appeared to be on the way out, pushed aside by the more glamorous discipline of molecular biology. Government funding for this field, widely perceived as stodgy, lapsed in the 1980s, and at universities across the country, taxonomists lost office space, positions, and respect.

Now this discipline is remaking itself into a more rigorous, hypothesis-driven science. Increasing numbers of systematics researchers have embraced molecular biology techniques and evolutionary principles. Universities are once again hiring them, and in 1995, the U.S. National Science Foundation (NSF) created a program designed to promote this resurgence. With about $4 million a year, Partnerships for Enhancing Expertise in Taxonomy (PEET, web.nhm.ukans.edu/peet) is creating a new generation of systematists comfortable with molecular and computer tools as well as with the microscope and collecting kit.

And just as critical, PEET awardees are also helping to spawn a new field: biodiversity informatics. PEET scientists and others are feverishly putting collections data online. Although daunting, the task is critical, says ecologist Jim Reichman, director of the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara, as these data sets provide a baseline for the next generation of biodiversity studies, in large part by providing a historical context. "Our hope is that the bioinformatics will give biodiversity a brighter future," says Patrick Crist, a conservation biologist with the U.S. Geological Survey in Moscow, Idaho.



Coexistence. This 2.8-mm beetle lives within army ant colonies.

From ledgers to keyboards

Together, the world's natural history museums house about 3 billion specimens, some accompanied by notes about how these organisms lived, reproduced—even what they ate. Collections can date back centuries, and paleontologic repositories provide a view of life going back millions of years (see sidebar). With these records, researchers can track changes in distribution through time and therefore assess the impact, say, of global change.

Distinguishing features. Scanning electron micrographs show details of mouthparts that can be used to identify beetle species.

But because these data are spread across many institutions, they have been notoriously hard to compile and use. A decade ago, anyone interested in learning about a particular species had to scan index cards and ledger books in which numerous biologists had recorded what they had collected and where. To see what was contained in other museums, the researcher actually had to visit them. Eventually, say leaders in biodiversity informatics, that will no longer be necessary: Researchers will need only log onto a Web site that gives them inventories of what each museum contains, sortable by species, geographic location, and perhaps even habitat.

The University of Kansas Natural History Museum and Biodiversity Research Center is at the forefront of this new movement. "It has done a lot in terms of leading the way in the integration and coordination of collections" online, says Terry Gosliner, who studies mollusks at the California Academy of Sciences in San Francisco. Although the Kansas museum is a staid, limestone edifice built in 1901, behind the scenes, PEET scientists and other curators have broken out of the traditional mold, thanks in part to the vision of a new director, Leonard Krishtalka.

At the museum, a plant physiologist–turned–computer expert named Dave Vieglais, for instance, has amassed a veritable database empire, known as Species Analyst (habanero.nhm.ukans.edu), making possible one-stop shopping for specimens. Now, thanks to Vieglais and his colleagues, instead of paging through ancient ledgers a researcher can type an organism's name into a personal computer to find out not only what's in the Kansas collections but also whether samples of the species exist among the 12 million specimens at several other museums, a number that will soon increase to 38 million (*Science*, 7 May 1999, p. 888). "It's one of the more exciting projects that is coming along," notes Stanley Blum, a bioinformaticist at the California Academy of Sciences.

To pull it off, Krishtalka, Vieglais, and colleagues had to overcome some entrenched views. Historically, museums have tended to be possessive about their collections. "There's

CREDITS: (TOP) DOUG KOCH/KU UNIVERSITY RELATIONS; (BOTTOM) KU NATURAL HISTORY MUSEUM/BIODIVERSITY RESEARCH CENTER

Fossil Databases Move to the Web

Paleobiologists readily acknowledge that they lag behind disciplines such as molecular biology in sharing data on the Web. But several researchers are working to put existing data sets online. And at least one team hopes to build a sort of GenBank of paleobiology, a Web site where everyone can deposit their fossil finds. Organizers of these efforts face a big hurdle, however: deep divisions within the community over the sharing and quality of data.

Although a Web search will turn up at least a dozen paleo databases, many consist merely of photos of selected fossil specimens or taxonomic lists. A few broader databases limited to one region, animal or plant group, or time period can be downloaded from the Internet in one huge chunk. What's missing are comprehensive, open-access, interactive Web databases that archive published data on where and when a species lived, information that is critical for analyzing patterns of evolution and extinction.

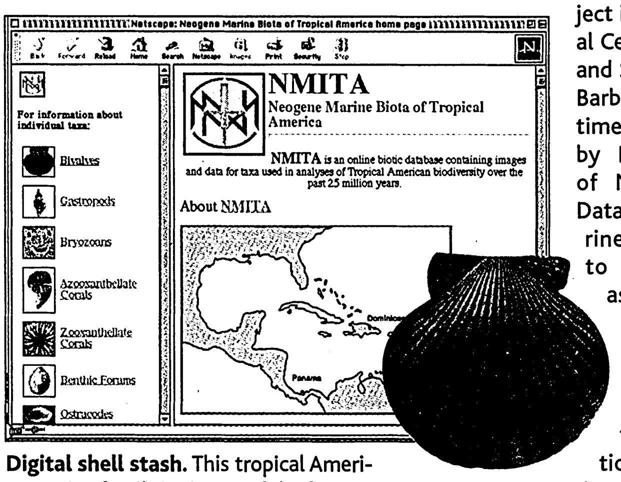
Paleobiologists point out that they face the difficult task of integrating species, temporal, spatial, and geochemical data that can quickly become obsolete if new fossils are discovered. "It's a much more complicated endeavor" than even living species databases, says Doug Erwin, paleobiology curator at the Smithsonian's National Museum of Natural History (NMNH) (see main text).

But would-be Web database builders must also deal with an ambivalence over such repositories that goes back to the most famous one, a marine species database built in the 1970s that the late Jack Sepkoski of the University of Chicago used to overturn many ideas about extinctions and diversification. Besides logging his own fossil finds, "he grazed and browsed in the literature and used it in ways that made some paleontologists uncomfortable," says Kay Behrensmeyer of NMNH. Such attitudes are still common among vertebrate paleontologists, whose fossils are relatively rare. These views play out both as reluctance by some collectors to share their specimen databases and as long-running disputes over the quality of compilations such as Sepkoski's—whether he used correct taxonomy, for example.

Despite that baggage, a few broad Web paleodatabases are under construction. They've been spurred by advances in "relational database" software that make it possible to dovetail separate data sets so that, say, shifting the Eocene period by 1 million years doesn't mean having to adjust every entry in a database. Such tools "made a huge difference," says Charles Marshall of Harvard University.

One impressive experiment in cooperative database building is Neogene Marine Biota of Tropical America, which logs marine fossils from the last 25 million years (porites.geology.uiowa.edu/index.htm). At this stage, the site, which emphasizes photos, is basically "like a Peterson's field guide" for identifying specimens, says co-curator Ann Budd of the University of Iowa in Iowa City. But her group plans to add data that's now accessible only to contributors so that other users can plot ranges and evolutionary trees.

Another project just coming online is the Evolution of Terrestrial Ecosystems (ETE) database developed by NMNH and John Damuth of the University of California, Santa Barbara. It covers both animal and plant terrestrial fossils and includes age, species lists, body size, and diet for nearly 4000 localities, largely from the African late Cenozoic. This week, ETE debuted a pilot Web version of the database (etedata.si.edu). Behrensmeyer hopes more researchers will contribute. "Once people see what can be done, I really think they will be willing to provide access" to their data, she says.



Digital shell stash. This tropical America marine fossil site is one of the first interactive paleobiology Web databases.

An even more ambitious project is under way at the National Center for Ecological Analysis and Synthesis (NCEAS) in Santa Barbara: The goal is to span all time periods and organisms. Led by Marshall and John Alroy of NCEAS, the Paleobiology Database is starting with marine paleofauna but plans to fold in other data such as Alroy's own North American mammalian databases (www.nceas.ucsb.edu/public/pmpd). "We need to have integrated databases to answer big-picture questions" about evolution, says Alroy. An open database where anyone can enter data via the Web "represents the only rational solution," he says. Alroy has found 36 collaborators so far but predicts that some private databases "will never be online."

Indeed, skeptics of the all-in-one database idea abound. Richard Stucky of the Denver Museum of Nature and Science, who's compiling Cenozoic North American mammal data to expand an older database called FAUNMAP, asserts that he's painstakingly gathered data using "strict criteria" to address specific research problems. Asks Stucky: "Can a central database answer all the questions [researchers are] asking? I say, 'No.' "

But others look forward to a day when anyone can troll a central fossil database. Says Marshall, "It seems daft to go into the field and collect a bucketful of fossils and not see it recorded anywhere."

—JOCELYN KAISER

a very fierce streak of independence," notes Kansas informatics expert Jim Beach. Data-sharing is often carefully worked out among researchers studying a particular group of organisms. And often individuals want to wait until they've published all their results before they open their books to colleagues. Complicating matters further, each museum has had its own way of doing things: Some, like Kansas, compiled information in ledgers, others used computer text files, and still others developed spreadsheets.

To Vieglais fell the task of coming up with a system that could work with all these types of collections. He first tested his software in 1998 by incorporating collections from the Kansas herbarium along with the museum's bird and mammal data into the prototype Species Analyst. It was able to retrieve information, despite differences in the formats of the collections. Soon afterward, he added mammals on file at the University of California, Berkeley, and he's been signing up museums, collection by collection, ever since. He lets the curators in charge of these collections decide how much to put online, so they are comfortable with sharing their hard-earned data.

As an added feature, Species Analyst forwards the geographic information about a given species in each collection to the San Diego Supercomputer Center. There, a program called GARP developed by David Stockwell maps that information and, based on the environmental data available for those sites, predicts the species' environmental niche and its overall distribution. For instance, after the Asian long-horned beetle was discovered in Chicago and a few other sites in the eastern United States in 1998, Vieglais, Kansas bird curator Town Peterson, and their colleagues used their program to determine where it might choose to live in the United States. Their preliminary run was "encouraging," says Vieglais, in that it shows the program works, "but discouraging in that it identifies most of the eastern United States as potential habitat for this invertebrate."

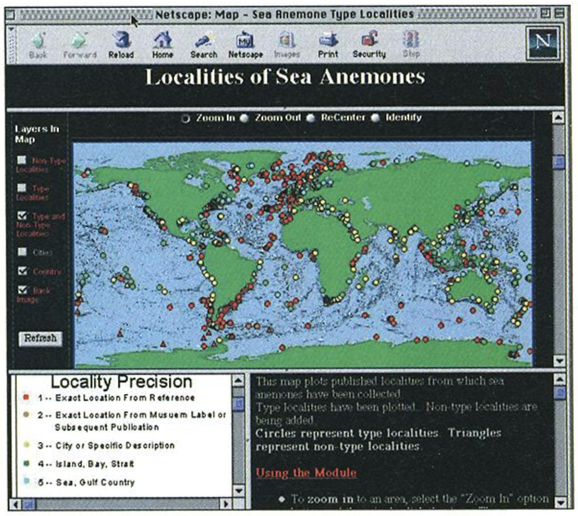
At this point, Species Analyst will provide the distributions of a particular species and report those results back to researchers at their own computers. But Beach is already working on the next generation of bioinformatics software, one that will create an archive of distributions for all the species contained within the Species Analyst fold. This archive will enable researchers to do more computer-based assessments of the organisms under study. With a $2 million NSF grant, Beach and colleagues from California, New Mexico, and Massachusetts will spend the next 2 years creating this archive. Once it is in place, Beach hopes to set up a Web site that will enable users to ask for a list of species that inhabit any place on the globe. In addition, researchers will be able to look for overlap in the habitats of, say, sister species, or of a plant and its insect pests.

Working on the Web

As bioinformaticists, Beach and Vieglais appreciate how computers can revive interest in and the use of taxonomic collections. With impetus from their PEET grants, some traditional systematicists are learning this same lesson. The PEET awards, started in 1995, come with a key stipulation. To receive one of these 5-year grants, researchers must agree that in addition to classifying little-known species and preparing the next generation of taxonomists, they must develop Web-based resources describing their organisms; the Internet, NSF's board realized, can open the museum drawers and cabinets to a much broader constituency than the few researchers who have access now.

Entomologist James Stephen Ashe, who runs one PEET team at Kansas, is attempting to sort out a large group of tiny beetles, the Aleocharinaethe—the type of work he has always done, but with a new twist. "There are so many and they are so small that they have overwhelmed the taxonomic expertise," Ashe says. Even beetle experts often can't pin down the genus, let alone the species, of these beetles.

To aid in this endeavor, Ashe, a postdoctoral fellow, and two graduate students have been assessing the evolutionary relationships of the oldest Aleocharinae species. They have already drawn up a list of names and synonyms of these beetles and are working out identification guides to the Aleocharinae in North America and Mexico. And rather than publish these in some obscure journal, they are putting their work online. They are also posting illustrations to help biologists figure out what they've got. To date, the Web site contains 1300 pictures, drawings, and scanning electron micrographs of body parts from 350 genera. It's possible to call up a set of images by the genus name, say, or call up legs from an entire group.



Pinpointed. Web maps now show where researchers have found sea anemones, including this bizarre corallimorpharian *(left)*.

Kansas marine biologist Daphne Fautin, one of only four sea anemone systematists in the world, also sees illustrations as critical for the proper identification of her group of organisms. She received a PEET grant to build a global database of sea anemones. Early in the project she realized few people had access to the original literature in which many sea anemones were first described. For that reason, her Web page includes historic illustrations from some century-old reports.

Fautin's goal is to sort out just how many species there are. Sometimes, a species was named more than once by researchers who discovered it on different sides of the globe. That members of the same species can look quite different only complicates the matter. Explains Fautin: "It's a big bookkeeping problem, and also a biodiversity and biogeographical problem." The 1300 species could really represent only 800.

To figure this out, Fautin tracked down the specimens that were used in the original description of a species. Now she has an inventory of what museums have which species, a resource that enables researchers at their office computers to easily identify the location of the specimens they are interested in. Unlike that in Species Analyst, her data set already covers all of the relevant museums, but only for sea anemones.

Early days

Despite these promising advances, biodiversity bioinformatics programs still have a way to go. Converting collection information into a digital form is tremendously hard—and boring—work. The data on the millions of specimens at each major museum could take tens of person-years to enter by hand into a computer, even if those institutions had the funds to do this work. So while fish collections are in good shape—a database containing the estimated 25,000 fish species was recently completed—many insect collections remain untouched.

Adding to the tediousness of the chore are entries about where the plant, butterfly, or elephant was found. Saying that a specimen comes from "5 miles [8 km] south of the intersection of Route 5 and Interstate 80" means little to a researcher in a different state, much less another country. For that reason, geographic analysis programs require locations demarked by latitude and longitude. So for the time being, someone must figure out from maps what the corresponding coordinates should be. "Just documenting what we already know is a tremendous challenge," says Gosliner of California's Academy of Sciences.

Finally, Fautin and others wonder whether technology will change so that 50 years from now, Web-based taxonomic resources will be inaccessible. She, for one, still plans to publish a print catalog about her sea anemone work. But others, such as Blum, think those worries are unfounded. The future is in the Web, he insists, citing a dozen projects like Species Analyst as just the tip of the electronic taxonomic iceberg. "We need to stop committing the information to paper," he says. His attitude, and that of an increasing number of systematists, curators, and biologists, is gradually changing the face of systematics, says Beach, and making this "a tremendously exciting time."

–ELIZABETH PENNISI

CREDIT: GEORGE MILLER