

All living things, online

Can taxonomy shed its dusty image and reinvent itself as a vibrant discipline for the Internet age? Virginia Gewin talks to the pioneers who are trying to turn this vision into reality.

Taxonomy has an image problem. Perennially short of funding, its greying practitioners are seen as dry and dusty — just like the museums and herbaria in which many of them work. “People who do taxonomy are depicted as postage-stamp collectors,” says Patrick Kociolek, executive director of the California Academy of Sciences in San Francisco, and an authority on the taxonomy of diatoms — algae that secrete cell walls made from glass.

Repeated reports from concerned organizations have echoed Kociolek’s observation, describing a discipline in seemingly inexorable decline. But hope may be at hand, in the form of initiatives that aim to reinvent taxonomy for the online generation. One project, called the Catalogue of Life, is taking firm steps to establish a federation of interoperable databases documenting the world’s taxonomic knowledge. Another group, the All Species Foundation, is even talking about naming and describing all living species within a single human generation.

That would be an incredible feat. There is currently no consensus on the total number of species: estimates range from 4 million to 100 million. In the 250 years since the great Swedish naturalist Linnaeus invented the modern taxonomic framework, roughly



Web of life: initiatives such as the All Species Foundation and Species 2000 aim to haul taxonomic information out of museums and herbaria to make it more readily accessible on the Internet.

hopes to quadruple the number of species described each year — a figure that currently stands at about 15,000. Today, there are at most 10,000 taxonomists worldwide, few of whom are in the developing countries that contain most of the Earth’s biodiversity. All Species aims to double the taxonomic expertise in these countries by recruiting ‘parataxonomists’ trained to collect, sort and document specimens.

Everything counts

Janzen helped to pioneer this concept through his involvement with the first all-taxa biodiversity inventory (ATBI) at the Guanacaste Conservation Area in Costa Rica. An ATBI is an audit of the entire biodiversity within up to 100,000 hectares of an ecosystem. Only three have ever been started — the others are in Hawaii and the Great Smoky Mountains National Park in Tennessee — and the Costa Rican project eventually dissolved when it lost local political support. One early goal for the All Species Foundation is to help to complete the Hawaii ATBI within five years. It is also in discussions with the US National Park Service about helping with further inventories.

In parallel, the foundation wants to identify and promote technologies that will speed progress towards its eventual target. It is evaluating tools ranging from pattern-recognition software for quickly identifying species from digitized images, through

1.7 million organisms have been named and formally described. And that information languishes in myriad journal articles, index cards, ancient books and museum collections.

Launched at a meeting at the California Academy of Sciences in September 2000, the All Species Foundation is an unusual alliance between leading taxonomists and some of the Bay Area’s high-tech movers and shakers, enthused by the prospect of a web-based project with an environmental flavour. The latter include Stewart Brand, who established the online community known as the WELL, his wife Ryan Phelan, who launched the consumer health website Direct Medical Knowledge, and Kevin Kelly, a founder and now editor-at-large of *Wired* magazine. “All Species is bringing new blood and energy to the field,” observes Dan Janzen, a biodiversity researcher at the University of Pennsylvania in Philadelphia.

“In all of my years, I’ve never met these kinds of people before,” says Terry Erwin, a taxonomist who specializes in carabid beetles at the Smithsonian Institution in Washington DC, and a member of the All Species Foundation’s governing board. “It’s really cool. I am usually surrounded by 12 million beetles, but I’ve gotten outside and talked with intellectual millionaires.”

To fulfil its mission, the foundation



Edward O. Wilson predicts that technology will rapidly accelerate the description of species.

HARVARD UNIV.

equipment to allow rapid DNA sequencing of informative portions of a specimen's genome, to hand-held interactive versions of the 'keys' that taxonomists compile to allow naturalists to determine which species a specimen belongs to.

"Roughly 10% of species have been discovered and classified in the past 250 years," says Edward O. Wilson of Harvard University, a pioneer of biodiversity studies who favours the lower estimates for total species number, and a member of All Species' science board. "With these technological advances, we can complete the remaining 90% in one-tenth the time."

All Species also wants to encourage the creation of digital images of at least half of the type specimens — or 'holotypes' — from which each known species was originally described. The resulting 'e-types', as they are being called, would then be easily accessible on the Internet, rather than sitting in the world's leading museums and herbaria.

Index links

But the All Species Foundation isn't the only game in town. Although most taxonomists applaud the foundation's audacity, many regard its aim of documenting all living species within a generation as unrealistic. Perhaps more attainable is the goal of the Catalogue of Life, a collaboration forged in June 2001 between Species 2000 — based in Reading, UK, and Tsukuba, Japan — and the Integrated Taxonomic Information System (ITIS), based in Washington DC.



Frank Bisby sees a consortium as the best approach for taxonomy.

Within 10 years, the Catalogue of Life aims to create a federation of databases that will together describe all of the species currently known to science. Its present compilation of 18 databases includes specialist initiatives such as Antbase, FishBase, AmphibiaWeb and ILDIS, a legume-taxonomy database started by Frank Bisby of the University of Reading, who is chair of Species 2000.

The scientists behind the Catalogue of Life envision a total of up to 200 component databases. Achieving interoperability among them has the potential to be a logistical nightmare, but Species 2000 has defined a common data model for member databases to adopt. Another issue is how to deal with the fact that many species have been given more than one name. For the 260,000 species in the Catalogue of Life so far, there are 420,000 synonyms to contend with. The Species 2000 team is researching how to create a tool that allows users to be referred automatically to the correct species informa-

tion from the various alternative names and classifications.

Having joined forces to create the Catalogue of Life, the leaders of Species 2000 and ITIS are now talking with other major taxonomic players about building the project into a wider consortium. In March, representatives of Species 2000 and ITIS met in Sydney and Canberra, Australia, with organizations including the All Species Foundation and the Global Biodiversity Information Facility (GBIF), a project conceived within the Organisation for Economic Co-operation and Development's Megascience Forum.

The GBIF aims to make a range of databases on biodiversity available to researchers and policy-makers. Following the Australian meeting, its leaders are currently considering a proposal to adopt the Catalogue of Life as one of its component projects — a decision is expected in October, when the GBIF's governing board meets in San José, Costa Rica.

The All Species Foundation, meanwhile, sees its role within the emerging consortium as a driver for technology development, and in creating a web portal that will provide links to other projects' data. "It will point to all the species-level information out there," says Phelan, who is All Species' chief executive. The foundation also recently unveiled a search engine that can return taxonomic information on about 1 million species already held in web-based databases.

Bisby points out that the consortium approach worked well for another recent 'big science' biology initiative. "We are taking as our model the Human Genome Project," he says. Both the genome project and the Catalogue of Life, notes Bisby, are truly international endeavours that aim to provide a vital information resource for a diversity of researchers.

Digging deep

Working in a consortium also has the advantage of eliminating potential duplications of effort, so ensuring that each project's funds are used to the best possible effect. The GBIF currently has a total operating budget of some \$3 million per year, and the Catalogue of Life has secured a similar level of funding in grants from US, European and Japanese agencies.

The All Species Foundation, meanwhile, has high hopes of winning multimillion-dollar funding from wealthy donors. But unfortunately, the current harsh financial climate is hampering All Species' fundraising efforts.

As the various projects strive to get their funding in place, some researchers are debating whether web-based initiatives such as the



Keeping track: Dan Janzen (above) recruited parataxonomists to assess Costa Rica's biodiversity.



Catalogue of Life could become the 'official' repository of taxonomic information — currently, original species descriptions are only recognized if they appear in print (see *Nature* 417, 17–19 & 573; 2002, and Correspondence on page 367 of this issue). Two-and-a-half centuries of tradition seem unlikely to be overturned overnight. But regardless of how this debate develops, the pioneers who are trying to reinvent taxonomy for the information age hope that, at the very least, they can help to dispel the dusty image that has done much to deter bright young biologists from entering the field.

"Old-fashioned taxonomy is less attractive to mainstream students coming into biology," says Bisby. "But we are revolutionizing the field." ■

Virginia Gewin recently completed an internship in *Nature's* Washington office.

All Species Foundation

▶ www.all-species.org

Species 2000

▶ www.sp2000.org

Integrated Taxonomic Information System

▶ www.itis.usda.gov

Global Biodiversity Information Facility

▶ www.gbif.org