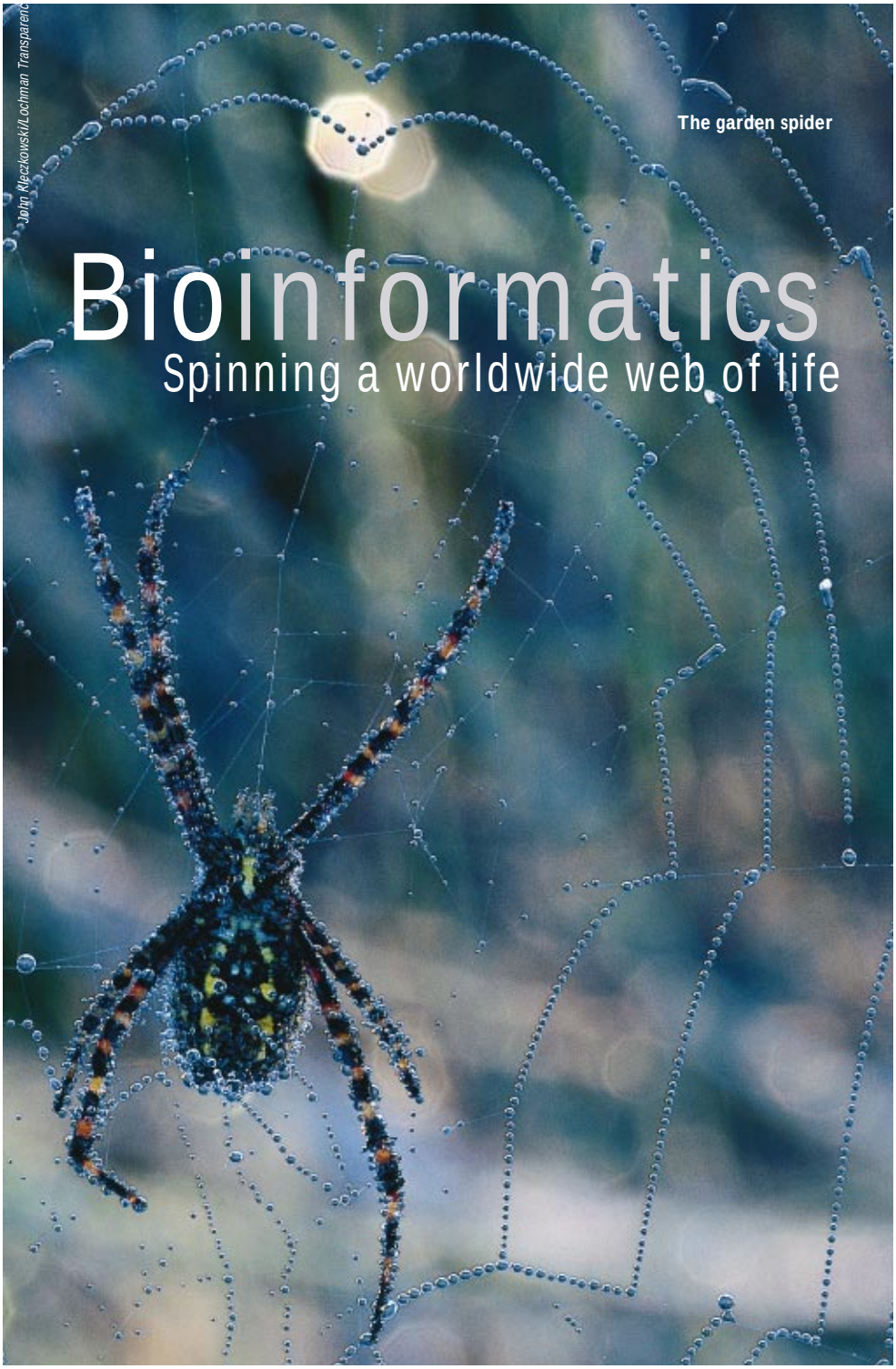
Bioinformatics

Spinning a worldwide web of life



The garden spider

Tim Thwaites reports on plans to dust off the data dinosaurs that lurk in biological collections worldwide.
Using advances in computer technology, the information can be made accessible and easier to analyse, broadening the knowledge base upon which decisions about conservation are made.

Students from schools in South Australia's Riverland region are spending time each month clearing insect traps in the nearby Bookmark Biosphere Reserve. They sort their catch into broad groups and send them off to CSIRO Entomology in Canberra for finer identification.

The primary and secondary students are involved in a long-term project to rehabilitate the reserve by planting native species and clearing weeds. By tracking the variation in species and abundance of insects over time, they can monitor the environmental changes they engineer.

It's important work in which a significant part of the students' science curriculum and outdoor education is devoted to renewing part of the Murray Basin. Added to this, the students will be at the forefront of a concerted thrust to gather and open up access to biological data through the use of computers and the World Wide Web.

CSIRO coordinator of the Bookmark project, entomologist Dr Geoff Clarke, says the students will eventually do almost all the insect identification themselves, using information, illustrations and software lodged on the Web. And the results of their work will be freely available. It's all being financed by the Federal Government through its Natural Heritage Trust Fund.

The concept of such a high level of community involvement in gathering biological data and contributing to conservation is so new that some of the computer tools to make it possible have not been developed. But once the project is up and running, Clarke hopes to replicate it across Australia.

'We hope to develop it as a module,' he says. 'Then we could transfer the idea to the arid zone or to tropical North Queensland. In this way, scientists and local communities can combine in meaningful conservation. If gathering this sort of data were left to scientists from Canberra, it would be enormously costly.'

The Bookmark project is at the forefront of a revolution known as biological informatics: employing computers to gather, store, combine, search, analyse, present and apply biological information. Advances in computing and the establishment of the Web have made it possible to pull together and analyse information from different disciplines stored in many parts of the globe.

Australia, particularly the CSIRO, is a world leader in providing ways to achieve this. Which is just as well. Because the developed nations of the world have grand plans for biological informatics. And they could well be realised using CSIRO-developed software.

Last June in Paris the Megascience Forum of the Organisation for Economic Cooperation and Development – the association of the world's rich nations – gave preliminary endorsement to a US$300

10.1071_ISSN0311–4546EC97p30

Melanie Brehaut, The Murray Pioneer

Brad Collis

Above: Dr Geoff Clarke is working with students in South Australia's Riverland region to collect insects in the Bookmark Biosphere Reserve. Eventually, the students will identify the insects themselves, helped by scientists via the World Wide Web.

Left: Dr Ebbe Nielsen chaired the organising committee of the first world conference on Biological Informatics. He says a Global Biodiversity Information Facility is vital to biodiversity conservation.

million project to make the more than 350 years worth of biological information stored in the Earth's museums and research institutions available to all via the World Wide Web. (The Megascience Forum is an inter-governmental committee of the OECD which deals with science projects too large for any single nation to handle.)

The project, called the Global Biodiversity Information Facility (GBIF) by the working group which recommended it, is seen as an essential step for studying, conserving and utilising the world's biodiversity. And the task is urgent. According to Dr Peter Raven, Director of the Missouri Botanical Garden and an advisor to the US President on biodiversity, as many as a quarter of the Earth's species of plants, animals, fungi and micro-organisms may be extinct by 2025, and three-quarters either extinct or on the way to extinction by the end of next century (see story on page 33).

This loss is compounded by enormous ignorance. 'Imagine in Indonesia trying to convert some wretched, acidic swamp land to rice paddies,' Raven says. 'When we do that, we are dealing with (ecological) communities in which we really are aware of the existence of only probably one in 20 of the organisms. These organisms are responsible for the flow of energy through those ecosystems, for regulating the cycling of minerals . . . for making them work. It's a very low degree of knowledge.'

The OECD working group views GBIF as a means whereby the rich nations, which hold most of the world's biological information and museum specimens but are home to relatively few of the world's species, can provide useful resources to poorer nations, which are responsible for managing a much greater variety of organisms.

Such 'repatriation' of information is already happening. Most type specimens of the birds of Mexico, and the accompanying descriptive and ecological data, are sitting in museums in the US, Canada and Europe. The World Bank is funding a project to transfer that data to Mexico. And the relevant Mexican authorities are using the information to plan conservation areas to secure the future of their birds.

Closer to home, the World Bank is also supporting a team of Australian ecologists, led by Dr Chris Margules of CSIRO Wildlife and Ecology, which is using informatics tools to recommend to the Government of Papua New Guinea which areas of vegetation are important to save from logging to maintain the country's biodiversity.

The backbone of the GBIF plan – which its proposers intend should be in operation in 2000 – is a definitive list of the world's 1.5 million named species. (No such list exists, even though named species represent less than one quarter of all organisms.) Eventually, each species name in this electronic 'catalogue of life' will lead directly to a description, references in the scientific literature, and information on where museum type-specimens can be found and where the organism is distributed geographically. The catalogue would also act as a gateway to further information on genetics, biochemistry, physiology, ecology, climate, medicine, agriculture – in fact, to any other scientific data relevant to the species.

CSIRO has a similar project – the Bioinformatics Initiative. This project is developing software tools that enable researchers to answer questions by gathering and analysing data on species held in CSIRO's plant, insect, wildlife and fish collections, and in other CSIRO databases. Links to the Australian Museum will also be tested. The program is coordinated by Dr John Curran of CSIRO Entomology and is being funded for the next two years by a grant of about $5 million from the Chief Executive Officer's Special Research Fund.

A related project, the Genetic Resource Initiative, will provide access to the latest computer tools for global retrieval and management of genetic data. And a third project, the Bioactive Molecule Discovery Initiative, will use computers to allow better screening and visualisation of data on molecules of interest to agriculture, drug development and nutrition.

Given this background of activity, it is no accident that the director of CSIRO's Australian National Insect Collection, Dr Ebbe Nielsen, was part of the working group which recommended GBIF to the Megascience Forum. This is the only way forward for biodiversity conservation, and it has to be done now, he says. Nielsen chaired the organising committee of the first world conference on Biological Informatics which was held in Canberra within 10 days of the Paris meeting which recommended GBIF.

One of the scientists attracted to that conference was Professor Meredith Lane, of the University of Kansas Natural History Museum. She is an adviser to the US

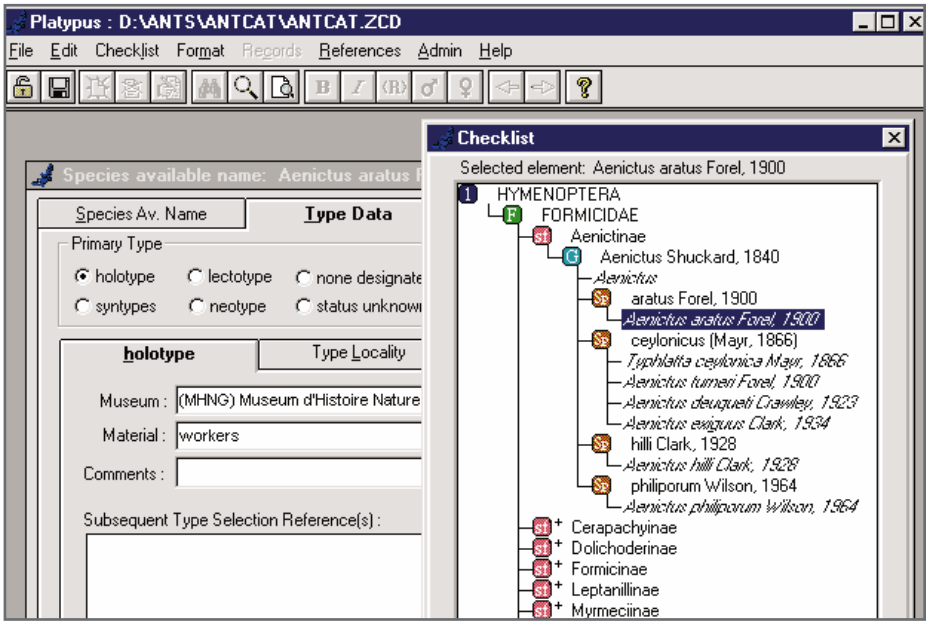
President on biodiversity and ecosystems and another member of the working group which recommended GBIF. 'Australia is far and away ahead of the rest of the world in expertise and knowledge in this area,' Lane says. 'It has a "just do it" attitude, while the rest of the world dithers. Australians are heavily relied on as the experts in the field.'

Lane says a typical example is eGaz, a program that converts the standard geographic locations written on museum labels – for example, 10 km north-east of Fiddlers Flat – into latitude and longitude. Such numerical geographic data can then be processed by computers for several uses, such as automatically mapping the distribution of species and species assemblages for research and management purposes. The eGaz (for electronic gazetteer) program, to be released by CSIRO Publishing next year, was written by Dr Steve Shattuck of CSIRO Entomology. It is the kind of program which will be necessary to convert the world's museum data into a useful electronic form.

'This is incredibly important software,' Lane says. 'And Steve produced it without much fanfare. He is an ant taxonomist, not a software person. But he had the support and resources of CSIRO to work beyond the bounds of his discipline to produce a generic program, which was related to the rest of his work.'

In fact, researchers across the breadth of CSIRO have developed or helped to develop a whole suite of software for linking and managing biological data. For instance, DELTA, a computer package to record descriptions of species and produce formal identification systems from them, was first written by Dr Mike Dallwitz, of Entomology, some 25 years ago, which makes it one of the oldest biological informatics programs in the world. A new version has just been released.

Then there is Platypus, a database program that collates, manages and reports taxonomic, bibliographic, geographic and geological information. It links references, type specimens, and distribution maps with species names – exactly the kind of strategy proposed by GBIF. Platypus was developed by a team led by Dr Keith Houston at the Australian Biological Resources Study (ABRS) – a federally-funded program of Environment Australia – and is used by organisations worldwide, including London's Natural History Museum and the Smithsonian Institution at Washington DC. Using Platypus, 25% of Australia's known fauna has been included in the Zoological Catalogue of Australia database, and another 10% is in preparation.



Above: The database program Platypus will be a key component of BioLink, a new biodiversity information management package.

Below left: Alistair Graham manages the Fish Collection at CSIRO's Hobart Marine Laboratories. The collection, begun in the 1930s, lends items to researchers in many parts of the world. A computerised index has been developed in preparation for future linking with other museum systems.

Below: Core information for taxonomic, genetic, biogeographical and ecological research on Australian wildlife is held in the Australian National Wildlife Collection (ANWC). Earlier this year, CSIRO Wildlife and Ecology received a $450 000 bequest from Sydney artist Sonia Farley. This enabled the establishment of a trust fund to further the work of the ANWC. A foundation was established to administer the trust and raise the profile of the collection. Serving on the foundation are CSIRO chairman, Charles Allen, chairman and managing director of Du Pont, John Foote, and Myer Foundation member Lindy Hayward. Wildlife and Ecology chief, Dr Brian Walker, says biodiversity loss is Australia's top environmental challenge, and biological collections must be maintained and accessible.