

MEGA SCIENCE

# Network of life

Software permits the integration of data banks with information on plants, animals and microorganisms

RICARDO ZORZETTO

*Published in December 2002*

In the very near future, Internet users will be able to access an integrated network of data banks with information on the name, classification and distribution of thousands of species of plants, animals and natural microorganisms in the state of São Paulo – essential knowledge for the definition of strategies for the preservation of the local biodiversity. As yet without a defined name, this network will integrate the data of the Phanerogamous Flora Project of the State of São Paulo, which for nine years has been cataloguing 7,500 species of flowering plants, with those of the network Species Link, which brings together information on twelve herbarium and museum collections in the state of São Paulo, and microorganisms registered through the SinBiota, the Information System of the Biota-FAPESP Program, responsible for the survey of the biodiversity and natural resources of São Paulo. It will also be the first step that may allow the Brazilian participation in a much wider network, directed towards the conservation of the biological diversity of the planet.

Technically the fulfillment of this initiative, pioneer in Brazil and, in some aspects in the world, has become possible thanks to a protocol of communication named DiGIR (Distributed Generic Information Retrieval), which allows tracking down information in distinct data banks and presents the results to the user as if the information had come from a single data bank source. The in-

ternational team that developed the software just about a year ago presented the most recent tested version, the last one before DiGIR 1.0, at the end of October in the town of Indaiatuba, in the interior of São Paulo, during the forum entitled Tendencies and Development in Information Technology for Biodiversity – the branch of information technology directed towards the creation of tools applicable for the study of the distribution and analysis of species. The team includes researchers from the Reference Center on Environmental Information (Cria in the Portuguese Acronym), the institution responsible for the maintenance of SinBiota, and Australian and German specialists and others from the universities of Kansas and California, in the United States.

The idea of developing a program that would integrate data from different systems – the so-called inter-operability, essential for allowing access of a greatest number of researchers to the greatest possible quantity of information on biodiversity – came about at a meeting of the Work Group on Taxonomic Data Bases that occurred during 2000 in Frankfurt, Germany. “Up to that moment, each network had developed its own software for exchanging information. We were on the road to developing the Species Link, and probably we might have created yet another different program, without the capacity of being able to inter-communicate with the other networks”, explains the systems analyst Ricardo Scachetti Pereira,

from Cria. “During the meeting the idea of DiGIR was presented, which would allow the exchange of information. Consequently we decided to participate in the development of this protocol of communication.”

Exactly because of the possibility of exchanging information between different data banks, the influence of DiGIR must not be restricted to the São Paulo network. It will also be a tool to be used in the virtual integration of this network with other networks of data banks on biodiversity existing in North America (*Species Analyst*), in Europe (*European Natural History Specimen Information Network*) and in Australia (*Australia's Virtual Herbarium*). And furthermore, according to the researcher Vanderlei Perez Canhos, director president of Cria, the DiGIR program is even a strong candidate to be adopted as the protocol that would make the integration of information systems of a multinational initiative, much more ambitious, namely: the *Global Biodiversity Information Facility* (GBIF).

**Catalogue of life** - Established in 2001, under the auspices of the Mega Science Forum of OECD (Organization for Economic Cooperation and Development), the GBIF is an international ini-

Difficult access: drawing of the shrub *Physocalyx major*, part of the work *Nova Genera et Species Plantarum*, by the botanist Karl von Martius

*Feb 1861*



PHYSOCALYX major.

tiative responsible for the coordination of a network that aims to integrate, and to make available through the Internet, not only the biology collections of the whole world – it is calculated that museums, herbariums and other collections store three billion examples of all the organisms extinct or currently living –, but also an electronic list with the name and the taxonomic classification of 1.75 million species of plants, animals and microorganisms scientifically described, the so-called catalogue of life. All of this, with the objective of assisting towards the solution of a question as old as it is important: the loss of the planet's biodiversity. "This is a global crisis in which everyone loses. It'll only be slowed down if concrete measures are taken to preserve the environment, for which it is necessary to have a large portfolio of scientific and technological knowledge", Canhos comments.

**T**hough the degradation of the environment has been under discussion since the 70s, the prospects for reversing the losses in biodiversity are not exciting. The very author of the term biodiversity, the American biologist Edward Wilson, from Harvard University, alerts in his most recent book *The Future of Life*, about the need not only of making data on described species available, but of completing, over a twenty five-year period, this inventory of living species along with the close to eight million still to be described. And for a worrying motive: over the last century human activity has triggered mass extinction never before seen since the Mesozoic era, which may eliminate or endanger a quarter of the plants and animals during the next thirty years. In an article published during 2000 in the magazine *Science*, Wilson estimated that it would be necessary to work twenty years and have an investment of US\$ 5 billion to complete this mapping, done in a comparable manner to the Human Genome Project, which took some thirteen years to sequence man's genetic material.

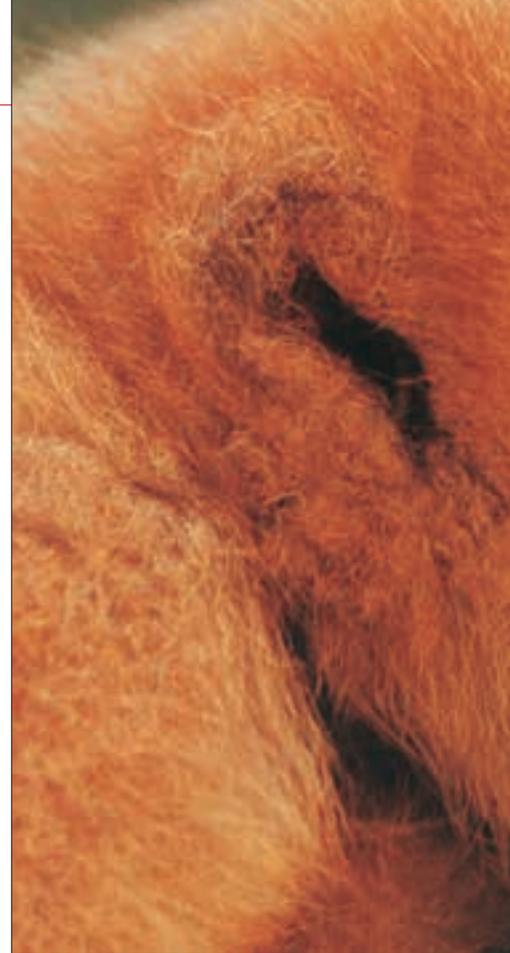
To turn the existing information available, to complete the mapping of the species and to analyze this data with information technology tools is an impossible task to be taken on by only a few institutions. It is an effort that de-

mands the help of researchers from several areas and the involvement of the greatest number of countries as possible, working in a cooperative and integrated manner – a characteristic of mega-science projects, kicked off in biology by the Human Genome Project. "No single country's capable of carrying this out alone", comments Dora Ann Lange Canhos, the project director at Cria. For this reason, at the OECD meeting during 1996, the establishing of a mega-structure (the GBIF) was recommended, capable of stimulating the participation of researchers and institutes of various countries – principally those that possess considerable biological diversity.

**Lack of consensus** - In spite of being the owner of the largest biodiversity on the planet, Brazil has still not officially joined the GBIF – which is made up of around thirty nations – and participates only informally, through the development of information technology tools for use in biodiversity. And it is not because of a question of budget – joining would cost the country around US\$ 50,000 annually, the equivalent of maintaining two doctorate students abroad for this period –, but because of a lack of consensus. While technical personnel from the Ministry of Science and Technology (MCT) and an expressive part of the scientific community are in favor of joining, the Ministry of Foreign Affairs believes that Brazil would be at a disadvantage, since it would cede more information abroad than it would receive in return – and, in this manner, the decision will remain with the future government.

"If we were officially part of GBIF, Brazil would only gain", says Cria's director. "We would be able to have more of an active participation in the GBIF's work program, as well as financing for the development of national projects linked to the repatriation of data on biodi-

Catalogue of life:  
digitalization preserves  
data about diverse  
species, such as the  
woolly spider monkey  
(*Brachyteles*  
*arachnoides*) and the  
solanaceas (right)



versity." In the opinion of Ione Egler, from the Secretary for Policies and Programs of the MCT, the admission of the country into the GBIF would be important mainly during this initial phase in which the members taking part in the initiative are defining action priorities. During the forum at Indaítuba, organized by the Cria, with financial support from FAPESP and other institutions, criteria were established for the selection of projects involving the digitalization of data from biological collecting, which should go into operation during the first semester of 2003.

The active participation of Brazil in the implementation of the GBIF would contribute to clearing up one of the principal difficulties of national research in this area: the lack of access to information about Brazilian species that are found abroad, mainly in the North American and European institutions, in which there are deposited examples and Brazilian types (a type is an example used to describe a species) collected by historical exploration expeditions such as those of Hans Langsdorff and Karl Friedrich Philipp von Martius. Out of the GBIF, Brazil is lo-



PHOTOGRAPHS BY MIGUEL BOYAVAN



sing, for example, the opportunity to propose priority actions, such as the digitalization of information of national interest, which would permit to speed up the repatriation of data existing in institutions abroad. To illustrate the need for access to this information, Ione cited the case of the Phanerogamous Flora of São Paulo: "There are some 7,500 species identified in the flora of São Paulo, but less than 500 types can be found in national collections. The rest is abroad. With this material digitalized, we would cede information on about 500 species, but would receive information on about 7,000."

**Flora Brasiliensis** - Another example of this difficulty of access is that which occurs with the work named *Flora Brasiliensis*, a collection of forty volumes edited by von Martius with information about close to half of our national flora, today estimated at 56,000 species. Most of the work's copies, mainly those complete and well preserved, can only be found abroad. As a consequence, this resource is underused by Brazilian researchers, according to the botanist George Shepherd, from the São Paulo State University of Campinas – a situa-

tion that could begin to change shortly: Shepherd is coordinating a project, as yet in its initial phase, that aims to digitalize a complete copy of this work, an initiative that could serve as a base for the production of a new Brazilian flora – digital and updated.

**I**n an independent manner, foreign institutions have already been providing this type of access to other countries. An example is New York's Botanical Garden, which made available, on line, data about 600,000 examples of plants – from a collection of seven million. But as yet this is insufficient. To better understand national biodiversity, it is necessary to have access to information on the largest possible number of collections.

Ione Egler, from the MCT, was emphatic: "Without information about our species, the country will not be able to recognize its biodiversity and will not implement the Biological Diversity Convention (*which looks to promote conservation, the sustainable use of biological resources and the fair redistribution of the benefits coming from the use of genetic resources of this biodiversity*)". In the case of not integrating into the ini-

tiative, the country will fail to even benefit from the possibilities of improving the qualification of personnel which the GBIF promotes and will remain without an active role in the definition of the information system adopted by it.

A frequent doubt, when one is dealing with the availability of information about national biodiversity on the Internet, is to know if this will make biopiracy any easier. "This is an incorrect idea. In the case of the GBIF, we're not dealing with material leaving the country, but an exchange of information of which there is a lot available within scientific magazines", Ione explains. The form of how the system is being proposed even allows that the researcher controls the information that he intends to place on the network.

But the greatest difficulty that the team from Campinas envisions for the effective implementation of the São Paulo network is to change the mentality of researchers so that they begin to share data and to work in a cooperative manner, as the very project demands. "People need to perceive that scientific information is essential for the formulation of policies and for taking decisions", says Dora. •