

## BUILDING ON NANO-DIVERSITY

# Microbial Information Management in Thailand

► By CHEN NAN

Microbes play essential ecological roles in balancing nature. Genetic resources contained in microorganisms are largely still waiting to be tapped, although the biological activities and organic compounds they produce are widely used in agricultural activities, pharmaceutical industries, food production, environmental controls and restorations. The direct economic significance of microbes also provides strong arguments for influencing the decision-making process benefiting biodiversity conservation in general.

Although only a small fraction of microbes are named and documented, unlike taxonomical groups with low total species numbers but high public interests such as vertebrates or higher plants, huge amounts of data are being collected largely because of the high rate of new microbial strain discovery. Not all collected data can be published in traditional ways, and a large portion of collected data remains in private notes only. These data face the danger of being lost through the eventual retirement of the data owners. Although some seemingly insignificant data such as microhabitat records of a well known species, or the fact that a species shows negative results in anti disease tests, may not deserve a place in a scientific journal, the accumulation of such data across time and space will enable us to detect possible important patterns and trends. Sometimes records as important as possible new species discovery remain undocumented simply because the

hard evidence, the strain, dies out. Data conservation issues must be treated seriously.

Traditional vehicles for dissemination of research results are paper-based. Certain shortcomings can be identified. Papers can be considered as two-dimensional and data are usually listed to accommodate only one way of viewing. By far, the most common published species inventories are arranged by species names. While in using this kind of inventory, it is not difficult to find out the distribution, habitat uses, substrate, growth medium, preservation methods and other information for given species, it becomes hard to supply basic things like "A list of all species (or endangered species) that occupy certain habitat types". It becomes complicated to answer a question such as "What is the number of species that more than three communities claim to be effective for medicinal uses but which have not been biologically screened?" Microbial species inventories maintained on papers can help little in obtaining information about sites (i.e., for environment impact assessment). In short, obtaining a subset of records that satisfies a range of criteria is extremely impracticable or virtually impossible.

To work around this limitation, inefficient techniques are used. Many books with duplicated information have already been published. It is not surprising to see book titles like "Forest Fungi in Thailand", "Seed Fungi in Khao Yai National Park", "Poisonous Fungi in Thailand", "Edible Fungi in Thailand", and "Endangered Fungi of Thailand". These lists are merely subsets of a larger

data bank distinguished by certain attributes. Data redundancy created in this way makes data updates even more demanding and also results in data inconsistency.

Modern information technologies may help solve some of these problems. Many commercial database management systems and software application development tools enable us to create sophisticated information management systems that run on widely available computer networks or even on the fast growing Internet. The general objectives for biodiversity information system projects are: 1) building mechanisms for compiling existing data in standard formats; 2) supplying tools for basic researches and for detecting information gaps; 3) establishing channels for serving up information in useful formats. The main function of an information system is to process data into information that, in turn, becomes knowledge.

To successfully implement a biodiversity information system, computer software experts with in-depth knowledge of biodiversity issues are needed. Developing a biodiversity system is far more than just composing software codes; it takes careful conceptualization, requirements analysis, designing, construction and testing. Only the right combination of correctly selected technologies, competent specialists and sound management would ensure the delivery and production of flourishing biodiversity information systems.

A case study would help the readers to understand better the various components and issues concerning a microbial information management system.

The National Center for Genetic Engineering and Biotechnology (BIOTEC) is under The National Science and Technology Development Agency of Thailand. One major function of BIOTEC is to take advantage of Thailand's vast microbial natural resources. A field survey team collects various microbes from the wild lands and natural habitats. BIOTEC maintains a modern cultural collection of isolated microbial strains supplied by the field team. A fermentation laboratory grows selected strains in order to produce larger quantities of microbial extracts. These crude extracts are distributed to various screening labs to test their activities against certain diseases such as TB, malaria, cancer, HSV and HIV. Once a crude extract is found active, detailed biochemical experiments will be done in the bio-resource research facilities to pinpoint the exact compound. Simply put, the goal is to find organic compounds that have inhibition effects on certain diseases, while causing no or little adverse consequences to normal human body cells.

The entire process generates various categories of data that have complex interwoven relationships. Initially, different laboratories handled data separately. Sometimes, data were listed using low end, user-oriented database products, notably MS Access. Most of the time, it was even worse that data were handled using electronic spreadsheet products, mostly MS Excel. This arrangement or lack of arrangement causes certain deficiencies:

- **Lack of standardization.** Non-standard data are difficult or even impossible to analyze.
- **Non-existent security measures.** Some of the data are considered as classified information. However, data files have been copied as many times as people wished and passed around on the network.
- **Data redundancy and in-**

**consistency.** Since the same data has many copies, large data redundancy can be created. When data need to be updated such as classification revision, it is impossible to hunt down all the copies of the same item.

- **Inability to efficiently output data.** When data exists in different formats, in multiple copies and worse, in broken formats, it is hard to generate reports.

We eventually decided to develop a comprehensive microbial information management system. An 18-month in-house software development project started in 1997 and was completed in 1998. The result is the Microbe Management System (Microman) copyrighted by BIOTEC that has been in operation for about 30 months. The Microman system has already become the standard software for data inputting, data processing and information outputting covering BIOTEC's research activities, ranging from field specimen collection, lab culture collection, fermentation operation, and cytotoxicity to bioassay testing. Microman also handles data concerning locations, habitats, and indigenous knowledge of biodiversity uses as well as scientific literature. Microman covers the taxonomic groups of fungi, bacteria, yeast and algae. Currently supporting several laboratories and a couple of dozens of users, the system has gathered nearly one hundred thousand records of all kinds.

Microman is easy to use and it behaves and feels just like any other modern software application. An average user needs just a few hours' training to become productive. There are about 30 inputting forms in the system for accepting data. It also provides 10 kinds of standard reports such as species summary, species distribution map, locality summary, screening summary, researcher summary, catalogue of cultures and spe-

cies inventories. Users are assigned different roles and their access levels in the system are controlled following the center's policy. Usage of the system is automatically monitored. The database is located on a centralized server where users can insert new data, update existing records or query the database from workstations in their own laboratories. Users of the system would immediately feel any data editing by other users.

For readers who are technically comfortable, here is a technical summary. Microman is a LAN and ODBC based two-tier client/server system running under MS Windows. The front end was built mostly with the OOP data centric MS Visual FoxPro (VFP) development environment. Visual Basic for Application (VBA) is used to carry out automation tasks. MS Word, MS Excel and PhotoShop are automated for certain operations. MS SQLServer 7.0 is currently the back-end RDBMS. There are more than 40 tables involved in the system. An auto-documentation process indicates that the front-end application is comprised of nearly eighty thousand lines of codes. Microman is totally event-driven, highly modeless, and it has a Windows compliant interface. Microman is secure, supported by the MS SQL Server built-in security mechanism. We have stayed away from the Internet because of the sensitive nature of some data.

Although computer technology will evolve, biodiversity data maintained in databases must survive over time. Every piece of data in the system must be associated with the contributor for taking responsibilities as well as credits. One hundred or 200 years later, a database record may well become the only documentation about certain facts concerning biodiversity. ■

---

**Chen Nan** is a Technical Officer (Biodiversity Information Management) at the National Center for Genetic Engineering and Biotechnology in Bangkok, Thailand