GUEST EDITORIAL

The information age and agricultural entomology

Scott E. Miller

Bishop Museum, Box 19, 000-A, Honolulu, Hawaii, 96817 USA
e-mail: Scottm@bishop.bishop.hawaii.org

Prevention and resolution of agricultural entomology problems relies on current and accurate information, part of the overall information on biological diversity. The popular news media provide daily examples of how the information age is rapidly revolutionizing the way information is compiled, managed, and distributed. But what does the information age mean to agricultural entomology? What lies beyond CD-ROM? Within a few years, farmers in many countries will be able to call up a knowledge base on biological diversity, identify their pest organism, and see what is known about it. Using this same system, they could communicate with farmers in other regions who are struggling with the same pest, or interact with an international research team working on the problem. How will these farmers progress from having no access to telephones to having access to a biodiversity knowledge base? The technologies to provide these services are rapidly becoming available on a world scale. While these technologies present tremendous opportunities, they also present challenges to the traditional systems of compiling, archiving, and distributing information.

Problems

Some of the major problems preventing local identifications and systematics research in tropical countries include scarcity of correctly identified reference collections; scattered type and voucher specimens and taxonomic literature, not easily accessible because of physical location, lack of indexing, and language; or inadequacy of communication (Hawksworth & Ritchie, 1993; Miller, in press); and, in some countries, a lack of trained personnel.

Solutions

Emerging technology, such as distributed networks, can cut through some of these problems, and empower local users to resolve their own problems, as well as to facilitate training. Distributed networks provide each user with a fully functional system able to work independently and the ability to use data from other members of the network. For example, an end user could be simultaneously connected to specimen databases at several herbaria. The system can be designed to make the user feel that he/she is only using one database (e.g., the process of accessing multiple databases is transparent).

Internet, the concatenation of networks around the world using TCP/IP communications protocols, provides a vehicle for such services (Krol, 1992). Along with mail and file transfer, Internet allows fully interactive communications at great speed and accuracy. Where telecommunications infrastructure is not yet suitable for full Internet, other systems that do not provide interactive access (e.g., store and forward systems) are still useful.

There are over 1,500,000 host computers connected to Internet, serving an estimated 15 million individual users–and doubling annually. By mid-1993, 55 countries had opera-

Correspondence is encouraged and should be addressed to the Editors.

tional full Internet access, and over 120 countries had some form of international network connectivity (e.g., including BITNET, UUCP, and FIDONET). Many regional grassroots organizations have adopted variations of Internet technology, such as Pegasus in Australia and the Pacific, Alternex in Brazil, Afrinet in Africa.

Networking proponents point out that Internet allows the ultimate in democratization of information because it eliminates many traditional barriers to information access, especially time and distance. Critics worry that lack of technology will widen the gulf between the 'haves' and 'have nots'. Full discussion of this issue is beyond the scope of this paper, but several points are worth addressing.

Electronic messaging can be less expensive and more convenient than facsimile or telex wherever a computer and telephone lines are available. Improved error checking and correction routines in modern hardware and software are expanding the usability of existing telephone services, as well as reducing the time necessary for transmission. Studies in Latin America and eastern Europe have shown that e-mail services are at least one and often two orders of magnitude lower in unit cost than fax or telex, at similar levels of service (Ruth & Ronkin, 1992). The expansion of packet-radio systems provides great promise. A complete, solar-powered ground station (including microcomputer) for communicating with a satellite now costs about US$7,000 (Ruth & Ronkin, 1992). This technology is already being used to provide medical communications and information in sub-Saharan Africa. Worldwide satellite e-mail is now a focus for the communications and pager industry, so over the next few years the technology will advance rapidly, while the unit cost will continue to fall.

Providing an Internet link is far cheaper than building a library or a reference collection of insects. In many countries, it is cheaper than erecting a building or paying one salary for a year. In most countries certainly within a few years, the cost of computer technology will not be the limiting factor.

Emerging technologies in database management and expert systems, coupled with the power of Internet, have tremendous potential for use in agricultural identification and information services. Thompson *et al.* (1993) and others have already promoted biological diversity information bases combining relational databases, expert systems, and image processing. However, previously, these were 'finished' and published on CD-ROM, ending their continued development. I am suggesting an evolving, interactive system that reflects the continuing experience of systematists and users, made available through Internet. The applications that can be supported by a suitable information infrastructure are limited only by imagination and creativity. CD-ROM or electronic file transfer distribution could still be used for those without full Internet access.

To be most robust and useful, the biodiversity knowledge base will require cooperation, collaboration, and shared responsibility among the many constituencies providing and using this information. The quality and utility of a biodiversity information system depends on the taxonomic framework (names, relationships, and associated information) provided by systematic research, but the enterprise will involve a broad spectrum of participants contributing and using data. The latter should also allow systematists to focus on applying their training and resources to providing the systematic framework!

A distributed network approach includes the following features: 1. Reduction in effort—products created in one place can be used wherever they are needed. 2. Increase in availability—even an isolated individual in a developing country can connect to Internet (or at least to a network connected to Internet) with minimal delay. 3. Reduction in cost—less expensive than erecting buildings, and amassing libraries and collections. 4. Increase in accessibility—allows questions that are not supported by traditional ways of organizing information, encouraging cross-disciplinary inquiry.

Examples

The following applications are prime examples of what can be done with available technology. Library catalogue systems provide many examples of distributed networks (e.g., the CARL system in the United States). The SMASCH project, centered at the University of California, Berkeley, will link major herbaria in California, into a distributed network. Each herbarium will maintain its own database, but the databases can be used simultaneously. While many biodiversity database systems are only in development phases, the Australian Environmental Resources Information Network (ERIN) already includes over one million specimen records (mostly land cover plants), and operates on these data with a sophisticated suite of analysis tools including climatic modelling. Several interactive identification systems (multiple-entry keys) are coming into wide use, and identification packages are already available for grass genera, termites, beetle larvae, and fruit fly adults (Miller, in press). Based on a general format for recording taxonomic data, these packages also allow generation of natural-language descriptions, keys, diagnostic descriptions, and conversion of the data for use with phylogenetic analysis programs.

Data can be added to the information system in a modular way, as they accumulate from different kinds of projects. Because most pest species occur regionally, if not more widely, data generated for one place will likely be of interest elsewhere. For example, most of the major pest moths of Asia and the Pacific are included in the CABI project, Moths of Borneo. If the information in Moths of Borneo were translated into a database, complete with images and interactive keys, it would form the basis for identification of many of the world's major Lepidoptera pests. If these data were supplemented with information on pest species from other regions, say Kenya, Costa Rica, and Brazil, it would be an increasingly important world resource.

A core component of this information base will be specimen data from collections in museums and herbaria. These collections have already archived millions of specimens that include a tremendous amount of information on distributions in time and space, as well as natural history (Richardson, 1992). Experience with Australia's ERIN program has shown that it is far less expensive to extract pre-existing data from museums and herbaria than to recollect it in the field.

Democratization of information

The current trends in information management and distribution are just the beginning of tremendous changes under way. While it will take time for some of these changes to affect developing countries, the impacts are now being felt in many developed countries.

Online, electronic publication is emerging as a viable alternative to traditional, ink and paper publication. However, people like books, so some kinds of books will always be produced. But for the very expensive, data-intensive products of biological research, online electronic publication is a very logical and attractive alternative. Online publication allows instant publication, easy updating, inexpensive delivery, and selective but very wide distribution. To 'publish' means 'to make public'. Compare the impact of a document as a printed book lodged in 300 academic libraries versus being available to 15 million Internet users! While these numbers are impressive, the implications of online publication need to be considered at many levels, including staff promotion and granting schemes that demand traditional publications (Taubes, 1993).

The vast amount of information becoming available at low cost on Internet is also changing the economics and politics of information management. It is no longer viable to hoard information and try to sell it, because most of all of the information is available somewhere else at no charge. In recent years, some institutions have protected their specimen databases from public access in hope of selling the data. But many are now abandoning this strategy because: 1. It has proven impossible to recover the real cost of creating and maintaining the database in this way—you can sell some data, but not enough. 2. Other institutions are committed to making their data available as widely as possible at no cost. 3. There is great potential for conflict of interest, especially related to government grants and subsidies. In the USA and Australia especially, tremendous amounts of biological data are already available on Internet (e.g., complete authority files for vascular plants of the New World and Australia via the Biodiversity and Biological Collections. Gopher Server at Harvard University) and more is coming fast!

In the USA, there is great pressure to make data collected with public funds (including National Science Foundation research grants) available to the public. The Crown Jewels Campaign, a joint effort between Essential Information and Ralph Nader's Taxpayer Assets Project, is aimed at opening access to the federal government's most important information systems. Consider the impact on CABI's literature database sales if the databases from National Agricultural Library and National Library of Medicine became available for free on Internet!

Until early 1993, the public in the USA relied on the news media to tell us about the the activities of our government. Now, transcripts of White House press conferences are available instantly on Internet, as well as on other home computer networks such as CompuServe. Thus, traditional information filters are being removed.

If the middleman is being squeezed out of information distribution, how can we support the costs of producing and maintaining these vital information resources? I would focus on obtaining subsidies to support the enterprise, and selling the expertise to interpret the data, rather than trying to sell the actual data.

Challenges

The problem is not so much how to deliver the information, but how to produce quality data rapidly enough to fill the need! If funding were available for efficient translation of information from biological collections and to support systematists, a tremendous amount of information could be put to use in agriculture, forestry, conservation, and environmental management.

Issues of control and intellectual property rights must be addressed. A balance of centralization versus decentralization appropriate to needs and resources must be found. This balance will probably shift increasingly toward decentralization as users gain access and experience. Technology will not replace training and human resources development which will, of course, remain critical. New ways of collaborating and sharing information are evolving (Pool, 1993).

Conclusion

Several years ago, Internet was a somewhat exclusive club, with participation usually limited to those at universities and government laboratories. Through the dramatic changes mentioned here, Internet is now becoming a household word in many developed countries. Satellite and radio technologies will soon make Internet pervasive worldwide. Barriers to information access based on status and identity are being reduced, while Internet is becoming well enough known to be the subject of cartoons in popular magazines.

The growth of Internet and development of related technologies is happening whether we use it or not! We can help the technology to develop in a way that supports agricultural entomology (and biodiversity in general), or we can ignore it at our own loss.

References

Hawksworth, D.L. & Ritchie, J.M. (1993) Biodiversity and biosystematic priorities: microorganisms and invertebrates. 120pp. Wallingford, CAB International.

Krol, E. (1992) The whole Internet: user's guide & catalog. xxiv+376pp. Sebastopol, California, O'Reilly & Associates.

Miller, S.E. (in press) Development of world identification services: networking. *in* Hawksworth, D.L. (*ed.*) Identification and characterization of pest organisms. Wallingford, CAB International.

Pool, R. (1993) Beyond databases and e-mail. *Science* **261,** 841–843.

Richardson, B.J. (1992) The Australian Biological and Environmental Survey: role of museums. *Pacific Science Association Information Bulletin* **44,** 5–11.

Ruth, S.R. & Ronkin, R.R. (1992) Aiming for the elusive payoff of user networks: an NGO perspective. pp. 160–168 *in* Canhos, V. *et al.* (*eds*) Needs and specifications for a biodiversity information network: Proceedings of an International Workshop held at the Tropical Database, Campinas, Brazil, 26–31 July 1992. United Nations Environment Programme, Nairobi.

Taubes, G. (1993) Publication by electronic mail takes physics by storm. *Science* **259,** 1246–1248.

Thompson, F.C., Norrbom, A.L., Carroll, L.E. & White, I.M. (1993) The fruit fly biosystematic information data base. pp. 1–7 *in* Aluja, M. and Liedo, P. (*eds*), Fruit flies: biology and management. New York, Springer-Verlag.

Young, J.E. (1993) Global network: computers in a sustainable society. *Worldwatch Paper* **115,** 1–57.