Contributed Paper Session II-1

THE ROLE OF INFORMATION TECHNOLOGY IN PROVIDING ANIMAL HEALTH INFORMATION - A PUBLISHER'S POINT OF VIEW

Deborah L. Wiley

Andrea Powell

Marketing and Development Services
CAB International
Wallingford
Oxon
United Kingdom

ABSTRACT: CAB International has been providing animal health information in printed form for over 60 Years, and in various machine-readable forms for nearly 20 years. However, the tremendous advances in information technology over the past 5 years, and the expected continuation of this trend offer vast opportunities for information suppliers and users alike. This paper discusses relevant advances in technology and how they can best be utilized to provide new and better access to the literature relating to animal health. Some of the technologies covered include digital scanning and optical character recognition, magnetic and optical storage capacities, networking, CD-ROM, expert systems, and multi- media products. Examples are drawn from CAB International's existing range of products, and advances are viewed in relation to CABI's strategic plan for future developments.

CAB International (CABI), producer of the CAB ABSTRACTS database and the related series of abstract journals, has been providing animal health information in printed form for over 60 years, and in various machine-readable forms for nearly 20 years. Although the range of CABI abstract journals, including *Veterinary Bulletin, Index Veterinarius, Animal Breedng Abstracts* and others, to this day remains one of the most widely distributed formats of CABI information, recent years have shown a marked shift towards the use of electronic media, both for information dissemination and for information management by the end-user. Many CABI customers are finding that they prefer the flexibility and convenience of machine- readable information, especially if it can be used on their own PCs and on their own desktops. Not only does electronic data save them time, it also provides the opportunity to create and manage personal databases of information taken from several different sources. From the publisher's point of view, this flexibility also opens up many more possibilities for the distribution of data; no longer are we simply restricted to the printed medium or to the rigid structures of printed books or journals. Recent advances in information technology, such as CD-ROM, digital scanning, optical character recognition, networking, expert systems and multi- media, have enabled publishers such as ourselves to provide new and better access to the wealth of information relating to animal health, and to adapt it specifically to the individual needs of our users. This paper will explain how CABI has already harnessed many of these new technologies in order to meet those needs, and how projected advances are viewed in relation to CABI's strategic plan for future developments.

CAB International

It is necessary at this point to elaborate a little further on the status of CABI, since it is a unique and somewhat complex organization, with our approach to serving the information needs of the agricultural community, particularly in developing countries, needing to be reconciled with "commercial" activities and special finance considerations.

CABI is an international, intergovernmental, not-for-profit organization which is "owned" by its 29 Member Governments. These are principally countries which belong, or used to belong, to the British Commonwealth, and CABI had its origins as the Commonwealth Agricultural Bureaux. The organization gained international status in 1988, and is now open to membership by all nations. The newest member, Hungary, is the first non-Commonwealth state to join CABI. CABI exists to provide services to agricultural research internationally in

the fields of information, taxonomy and biological control. By virtue of its origins and its present membership, it has a special mission to serve scientists and the agricultural communities in developing countries. On the scientific side, this is done through the identification services provided by the Biosystematics Institutes, and the execution of biological control programmes funded in the main by the international development agencies. On the information side, the CAB ABSTRACTS database and the publications derived from it represent key reference sources to international agricultural research. These are widely used throughout the developed world, but less so in developing countries. There is a great need to improve the dissemination of information to scientists in the developing world, and modern Information Technology provides the opportunity to cut across the traditional barriers of cost and poor communications.

At the same time, CABI receives no government subsidies and relies heavily on the sales of its information products to support its scientific and information activities. This presents the need to develop and maintain successful markets for these products and to continue to develop and diversify the range of products to satisfy user demands. It is also necessary to continually monitor and, if possible, reduce the costs of production of the database in order to contain the costs of a process which is expensive because of the scientific expertise applied to abstracting and indexing. On both of these counts, Information Technology is intrinsically important to CABI and to the worldwide users of its services.

The IT strategy, for the production of the CAB ABSTRACTS database and also in our approach to CD-ROM publishing, is fundamental to CABI's viability. If it is successful, it will result in improved information transfer to the world agricultural research community, and will be particularly beneficial to scientists and students throughout the developing world. Reconciling the need for commercial viability with the objective of serving the developing world can be extremely difficult, particularly at times like the present when many publishers are under financial pressure. Yet it has to be done, and we firmly believe that new technologies, such as CD-ROM, provide important opportunities to facilitate the process.

Floppy Disk Products

Before moving on to illustrate how CD-ROM technology has revolutionized the dissemination of electronic data, I would first like to describe how we are making full use of more conventional technologies to satisfy the information needs of scientists and researchers involved in the field of animal health. More and more researchers are relying on personal computers for the storage and management of their information, whether this is in the form of bibliographic references, contact addresses of others working in the same field or their own notes accumulated over many years of study. PC networks enable several users within the same organization to access one central source of information virtually simultaneously, thus avoiding duplication of effort and multiple purchases of individual software programmes. Of course, computer networks bring with them their own problems and restrictions; networking software must be purchased and loaded onto each workstation on the network, networking licences must be obtained for every software program and for every CD-ROM database to be used, continuous technical support must be provided for the network and a certain amount of speed and efficiency may be sacrificed in favour of more widespread availability. Obviously, the advantages and disadvantages of establishing a local area network need to be weighed in each individual case and a decision taken accordingly. It is then up to the software producers and database producers, such as ourselves, to adapt their products to suit the needs of both single and multiple users and to make the position on networking unequivocally clear.

CABI's policy on networking CD-ROMs includes a surcharge for supplementary users within the same organization at the same geographic location, with dial-in access via a modem only permissible on negotiation. Any information from the CAB ABSTRACTS database supplied on floppy disk may also be used on a network or distributed by a subscriber to remote locations within the same organization, but with a few stipulations; any data sent to remote sites must be provided free of charge, since CABI is entitled to additional royalty payments on any income derived from our products; the CABI copyright statement must appear on each record distributed, and details of all recipients must be submitted to CABI for approval.

Returning to the move towards the use of PCs for the maintenance of individual databases, computer software producers have not been slow to recognize the need for sophisticated database management

programmes, which allow individuals, either on a local area network or on a stand-alone workstation, to edit and manipulate an amalgamation of data in a highly efficient and user-friendly manner. Data can be imported directly from external sources, such as online retrieval systems, CD-ROM databases or simply from other machine-readable data files, supplied on floppy disks. Once imported, this data can be searched, sorted and otherwise reprocessed to increase their intrinsic value in an almost limitless capacity.

CABI has followed this trend towards computerized personal database management by supplying selected sets of the CAB ABSTRACTS database on floppy disk, without any retrieval software, but in a choice of three formats which make them easily compatible with the vast majority of commercially available database management programmes. The comma-delimited format is recognized by most commercial programmes, while the ISO 2709 format has been specifically created for use with Micro CDS/ISIS software, which is distributed free of charge to developing countries by UNESCO. The third format is compatible with the popular database management programme, Pro-Cite, produced by Personal Bibliographic Software Incorporated.

Moving on to specific electronic products, it is now possible to receive all of the bibliographic citations and abstracts published in any one of the range of printed CABI abstract journals, such as *Veterinary Bulletin* or *Index Veterinarius*, on floppy disk. These electronic journals are provided at the same intervals as their printed equivalent, and, once loaded into a database management programme, can form the basis of a sophisticated personal reference bank. In order to maximize the use of these abstract journals on floppy disk, CABI is carrying out an ongoing evaluation process of as many commercially available database management programmes as possible, with the aim of compiling a list of recommended packages for use with our floppy disk products. The list includes such established programmes as HEADFAST (Head Software International), ideaList (Blackwell Scientific Publications Ltd), Reference Manager (Research Information Systems Inc), Library Master (Balboa Software), INMAGIC (Inmagic Inc) and Cardbox Plus (Business Simulations Ltd). Of course, the list is by no means exhaustive, and the evaluation process continues apace.

In addition to abstract journals, CABI is also able to provide a monthly current awareness service on floppy disk, with output based either on a user-defined search strategy or on a standard search profile determined by CABI scientists and information professionals. Many researchers and scientists have found this SDI service invaluable for remaining up-to-date with the latest published information in their particular field of study; it is impossible for anyone to find the time to read through every primary scientific journal, let alone the many non-serial publications issued every year. CABI's SDI service does this for them, and sends them the information of their choice in an easily manageable and compact form.
Floppy disks are without a doubt a very convenient way of disseminating information; selected sets of data can be delivered directly to the end-user's PC and, within copyright restriction, repackaged and reprocessed as required. However, there is still a limit to the amount of information which can be stored on one high density floppy disk, which is where the advances made by compact disc technology have been used to such tremendous effect.

CD-ROM Technology

It is well accepted that the advent of CD-ROM technology has had an enormous impact on the world of information dissemination and retrieval, particularly in countries where access to online databases has always been impossible because of unreliable telecommunication services and high costs. The advantages from the user's point of view are several and widely recognized, including the ease of retrieval, the non-reliance on telecommunications, the convenience of being able to access such a huge amount of data on one compact disc, and the portability and durability of the disc itself.

The benefits for the database producer are as many, but are looked at from a different point of view; the enormous storage capacity of a CD-ROM (650 megabytes) allows instant electronic access to a huge portion of the CAB ABSTRACTS database. In fact, each disc can hold over 400,000 individual records, which represents three years' input to the database as a whole, or the complete backfile in certain specific subject areas. Users of the CABCD (the entire CAB ABSTRACTS database on CD-ROM) will thus have simultaneous access to the complete range of CABI abstract journals rather than just selected issues or volumes, and their familiarity with CABI publications is bound to increase.

On the other hand, by limiting the subject coverage of a CD-ROM to a specific area such as veterinary science, animal production or forestry, many more years of data can be included, and the product can be aimed at a more specialized audience. Whilst the CABCD may not be appropriate for a veterinary

library, a CD-ROM focusing solely on veterinary science and animal health will be much more attractive. This ability to tailor the product to the needs of these individual markets enables us to satisfy their information needs much more efficiently and to provide them with an unparalleled archive of information at their fingertips.

VETCD and BEASTCD

It was in the fields of veterinary science and animal production that the first specialty CD-ROMs (collectively known as CAB SPECTRUM CD-ROMs) were conceived and produced. This was in response to very vocal demands from information professionals in the animal health and veterinary science communities. VETCD provides citations to over 450,000 source documents, including every record published in the core journals *Index Veterinarius* and *Veterinary Bulletin* since 1973. Also included are all of the citations and abstracts published in Helminthological Abstracts, Protozoological Abstracts, Review of Medical and Veterinary Entomology and Review of Medical and Veterinary Mycology, also back as far as 1973. In addition, a wide selection of records from other relevant abstract journals are included to provide comprehensive coverage. All aspects of veterinary medicine, arthropod, helminth and protozoal parasite infestation of wild and domestic animals, and fungal diseases are covered.

BEASTCD contains all records published in *Animal Breeding Abstracts, Nutrition Abstracts and Reviews - Series B: Livestock Feeds and Feeding* and relevant records from *Dairy Science Abstracts, Pig News and Information, Poultry Abstracts and AgBiotech News and Information*, also from 1973 to date. All records on dairy technology and breeding, feeding, management and biotechnology of economically valuable animals and of laboratory species relevant to agriculture are included. The year 1973 is a significant date because it was in this year that the CAB ABSTRACTS database first came into existence and that CABI began to gather together all of the records published in the various abstract journals to form a central computerized database. The three million records published by CABI before this date are still only available in print, making their inclusion on a CD-ROM that much more complicated. But there is a solution, and one which has already been utilized by CABI to create the third of the CAB SPECTRUM discs TREECD, which contains information on forestry, agroforestry and forest products from 1939 to date.

The technique employed to create TREECD was optical character recognition, or OCR. This involved the optical scanning of every single abstract published in Forestry Abstracts since its inception in 1939 and the conversion of each character into a machine-recognizable format for editing. Such a painstaking process was not without its problems, of course. Mathematical formulae and Latin or Greek characters were often incorrectly scanned, and extensive on-screen editing and the appending of index terms were needed to ensure the quality of the end product. Although the production of the disc eventually took longer than anticipated, the important first steps towards the use of OCR techniques have now been taken, and the way is now open for the same techniques to be used in other subject areas. Unfortunately, such a process is not only painstaking, but also rather expensive, and TREECD was produced in cooperation with the UK Overseas Development Administration and the Oxford Forestry Institute, without whom such comprehensive coverage would not have been possible. A similar procedure could be implemented for the older animal health information if a demand could be shown and funding obtained.

A second advantage of CD-ROM technology for the database producer is that its non-reliance on telecommunications makes it accessible to the large numbers of potential online searchers who are frustrated either by the cost or the lack of online facilities. As explained earlier, an area of particular importance for CABI is liaising with external funding agencies to facilitate the sponsorship of CD-ROM databases into developing countries, where access to electronic information has always been difficult or impracticable. In our experience, funding agencies have been quick to recognize the potential of CD-ROM technology for improving and updating local information services in developing countries, and the CAB ABSTRACTS database clearly lends itself well to such projects.

CD-ROM, for all its importance to our special fields of interest and activity, really represents a minor element of the new generation of information technology based around the microcomputer. In fact, in the computing sense there is nothing special about CD-ROM, given that it is simply an information storage and delivery medium, albeit of relatively large volume. Despite this, however, CD-ROM technology still offers some real new potentials in the provision of information, since it is the application of this technology which brings with it so many new possibilities.

One of the exciting new potentials offered by CD-ROM technology is the capability to integrate databases of differing kinds, and in the context of scientific research information, to cross the barrier between reference information and factual data. CD-ROM technology is not limited to the distribution of bibliographic databases like CAB ABSTRACTS; pictures, maps and charts can all be included on a standard CD ROM to form a multi media database.

CABI will be making an important move into the integration of the CAB ABSTRACTS bibliographic database with factual scientific databases or data collections of varying kinds. In certain disciplines, CABI is uniquely placed in terms of the information resources and scientific expertise available to us. The concept is to bring these resources together using information technology in order to provide new and comprehensive sources of information and data.
One such discipline, although this is only an example, is that of Pest Management, where CABI has special resources and capabilities, and the strategy can be illustrated by reference to a planned development in this area. This is to produce a multimedia database 'Compendium' comprising a number of databases stored on CD-ROM and linked and accessed using hypermedia technology. The 'core' is a full text database comprising detailed scientific descriptions of a number (several thousand) of pests and diseases of crops of economic significance worldwide. These descriptions will be compiled by scientists expert in the study of these pests and diseases which could include insects, fungi, weeds, parasites and other organisms. A similar compendium of animal pests and diseases could of course be produced using an identical pattern.

Associated with this core database will be: a comprehensive bibliographic database for these organisms selected from CAB ABSTRACTS; data defining the known worldwide geographic distribution of the pests, and able to be represented in the form of global or regional maps; illustrations, including line drawings and colour pictures, digitized for computer storage and presentation on graphics screens; reference databases providing such details as Natural Enemies, Biological Control measures, Taxonomic authority files; and so on.

The Compendium would thus represent a unique factual resource which could in fact only be produced and maintained using Information Technology. It also presents the opportunity to enhance the factual data further through the use of specialized PC software; for example, to assist in the identification of pests and diseases using taxonomic identification software, and (importantly) to utilize that factual data to perform predictive modelling of the potential spread of pests taking into account geographic and climatic factors. The same concept can be applied to diagnosing animal diseases, where expert systems lend a hand in identifying symptoms and ultimately problems. The diagnosis could then be linked with appropriate references (or maybe even full-text) suggesting treatments or therapy. This we see as the next generation of information delivery, where the power of the PC and associated technology provides for the first time the opportunity to cut across the traditional boundary between research information, as represented by bibliographic databases, and the huge store of factual data in the various forms of text, graphics and techniques for modelling and analysis. I'm not saying this will happen tomorrow, but it is certainly the direction in which we are heading.
The power of a PC frees us, the publisher, from the traditional bounds of information provision. As the sheer amount of literature available today threatens to overwhelm the researcher or information professional, we feel it is our responsibility and function to provide the appropriate amount and level of information to those who need it.