

PUBLISH YOUR DATABASE ON CD-ROM FOR PROFIT: INFORMATION FOR AFRICA

MARGARET CRAMPTON
National Inquiry Services Centre (NISC)
Grahamstown, SA

I. INTRODUCTION

This talk is directed at those of you with databases and those of who you are planning to produce databases. I wish to alert you to the potential for database publishing on CD-ROM as a means to publicise and make available your information resources beyond the borders of this country, at the same time earning revenue which will assist you with your information service.

I also address those of you with a need for bibliographic information. CD-ROM is a compelling medium which you can make available in your library, opening up an avenue to vast quantities of information in a very accessible way.

At this time of rising costs and static budgets many serials librarians face the dilemma of down-sizing their library. They nevertheless still need access to the information sources which they are compelled to cancel. An option to consider is the replacement of one or two journal subscriptions with a subject related CD-ROM title. A journal subscription can be substituted by a CD-ROM providing access to a library of information on one disc.

II. DATABASE PRODUCTION AND THE FISHLIT AND NISC ALLIANCE

In 1983 planning was initiated for a computerised bibliographic information system on fish, based on the extensive library of the JLB Smith Institute of Ichthyology at Rhodes University in Grahamstown. The FISHLIT database resulted from this initiative and the design, development and maintenance of this information system on fish is described by Crampton (1), (2). Funding to maintain this computerised database was an ongoing problem. The Institute

was compelled to derive revenue from the information service in order to ensure viability of the database. Publication on a NISC CD-ROM in 1993 was a major milestone.

In early 1995 FISHLIT was acquired by NISC SA (National Inquiry Services Centre Pty Ltd). Margaret Crampton is an Honorary Associate of the JLB Smith Institute as well as the Managing Director of NISC SA. In this way the close connection with the Institute and Rhodes University is maintained. FISHLIT, however, through joining NISC, will move from a national resource to become an international one as well, with a secure financial position.

NISC SA opened in Grahamstown, South Africa in early 1995. NISC SA will not only build databases but is a CD-ROM publisher. The staff at NISC actively seek out databases on all subjects relevant to Africa for publication on CD-ROM. These are marketed in Africa and worldwide with the CD-ROM discs derived from NISC branches in many countries. The potential exists for African information sources in the field of animal health and veterinary science to publish on CD-ROM.

III. THE COST OF DATABASE PRODUCTION AND INFORMATION SERVICE PROVISION

A. The Concept of Free Information

In the past there has been a belief that access to information must be free. In an ideal world this should be so, especially as the future progress of a nation is so dependent on a foundation of knowledge and experience. It is essential to progress without repeating previous research. The hard fact is, however, that the management

and provision of information is expensive even if the users perceive it as free!

B. The Cost of the Information Sources

The journals must be purchased and subscriptions to hundreds of journals is extremely costly. The annual subscriptions to the journals used for the FISHLIT database range from \$20.00 to \$9000 per title per annum. Although journals are also received as donations or exchanges, there is always a core of subscriptions which must be purchased.

In the present economic climate the annual cost of journal subscriptions is very difficult to predict and budget. It is affected by fluctuating exchange rates and rising prices. In general the costs rise far faster than the budgets of most academic libraries and these are the libraries with the potential to compile information databases. All over the world librarians are tackling the problem of down-sizing.

C. The Cost of Library and Information Centre Staffing

Staffing a database production centre is expensive and requires dedicated qualified staff who are prepared to work long hours and produce meticulous work. Needed are librarians to manage the document collection; indexers to select the articles and provide the abstracts and/or keywords; typists to do the data capture; information scientists to manage the information service and marketing staff to maximise the use of the database. Staffing is the most expensive aspect of database production and is ongoing.

D. The Cost of Computer Hardware and Programming

Database production requires powerful computers with large hard disc storage space, fast access to data and the ability to run multi-tasking at the required level without deteriorating speed. Such equipment tends to be expensive and requires

replacement at 3 or 4 year intervals to keep abreast with the technology.

Perhaps the most critical factor in database design and development is to work with programmers who are innovative, competent, and able to fulfil the needs of the database producers for efficient accurate data entry. The needs of the users for a friendly powerful interface with neat reports to capture references to disc or printer must also be provided for.

With the best database in the world, it is the programming which makes the information accessible in a useful fashion. With FISHLIT we have been fortunate to work with talented and experienced programmers from the departments of Information Systems and Computer Science at Rhodes University, and latterly the software development team of NISC. The local and American contributions have been invaluable.

IV ADVANTAGES OF CD-ROM PUBLICATION

Database producers have the ability to add to the library budget by earning revenue from the databases which they produce and there are several options one of which is to publish the database on a CD-ROM.

CD-ROM (Compact disc - Read Only Memory) technology allows rapid access to large and valuable collections of data. CD-ROM discs can span decades of research and compile data from different sources. The paper equivalent of each CD-ROM is about 400 000 pages. The high density storage of CD-ROM is about 1800 times that of a floppy diskette. The product combines the power of a Mainframe computer with the convenient access of a personal computer or the portability of a lap-top computer.

In Africa where the infrastructure is often lacking or rudimentary, where telecommunications are slow or unreliable and where LANs and WANs are not well developed access to databases on CD-ROM is still one of the best ways to provide information resources. CD-ROM discs are easy to send through the

mail and are robust and versatile and can be mounted on any platform, can stand alone or be networked.

V. COST RECOVERY THROUGH CD-ROM PUBLICATION

For the database producer CD-ROM publication provides an ideal source of extra income, particularly as there is no commitment of extra staffing needed. In 1990 the decision was taken to publish FISHLIT on CD-ROM and NISC (National Information Services Corporation) in Baltimore, Maryland was chosen as the publisher. NISC undertake to do any programming to customise the data and to master the discs without charge. They pay a royalty on every disc sold. It was felt that the ability to earn extra income without capital outlay was ideal.

VI NISC ALSO OFFERS SEVERAL OTHER UNIQUE ADVANTAGES

A Marketing

NISC has companies and agents worldwide. By joining an international company, with the ability to market aggressively in many countries, ones's database is assured of an extended market and the ability to reach a very much wider audience. This marketing effort is not an expense to the database producer.

Each of the databases on the disc has its own user base and each database is exposed to the user groups of the other databases. These databases are gleaned from all over the world and this provides another means by which one is able to reach a wider market.

As use of the CD-ROM is a lease arrangement and requires that the subscription be renewed annually the royalty income is ensured in perpetuity as long as the users renew their subscription. CD-ROM publishing ensures a predictable income as there are few cancellations.

B ROMWright software

The software development of ROMWright is done by the programming team at NISC. They have a reputation for being responsive to the needs of the customers and are making product improvements all the time. There is always context sensitive help and online help is so good that it is not often necessary to consult the manual if one reads one's screen. Peter Jacso (4) writes a regular CD-ROM column in the journal *Information Today* and in his annual assessment of the CD-ROM industry writes "CHEERS for the...Best non-multi-media company performance to National Information Services Corporation." He then goes on to enumerate the achievements of the ROMWright Software.

1. Search level

NISC provide three distinct search modes to suit each user's personal style: NOVICE provides a basic index, ADVANCED offers distinct-field searching, and EXPERT mode is an innovative, full-powered update of the traditional online format allowing set searching and the repeat of previous searches without tedious extra typing. It is dynamic and interactive and flexible, allowing previous searches to be kept and the software totals on screen as new searches are created. Searches can be transferred from the NOVICE to the ADVANCED to the EXPERT and are transformed to the new level automatically.

2. Assistance with choice of search terms

Databases are indexed in different ways and may favour terms not naturally chosen by the searcher. To avoid missing references through not using the correct term, NISC have some helpful features:

(a) The AUTODEX and ROTATED INDEX

These allow the user access to all the words used in the compilation of the records. The AUTODEX will pop up automatically and give the use frequency of particular single words. The AUTODEX often indicates the preferred version of a root term. Using the AUTODEX and the INS button new terms are added to the search strategy. The ROTATED INDEX is similar and will react to any word in the search strategy on pressing Shift F2. The rotated index is a permuted phrase index and gives access to the desired word when it is in a phrase. It can be useful to find other related uses of a particular word.

(b) P.I.C. Variant Searching

Plurals - Type the singular form and the plural is searched as well, automatically. Or type the plural and the search includes the singular form. There is no need to use truncation for plurals. The software will take care of such things as goose and geese as well as the regular plurals

International variants - International, English-language spelling differences and many similar variants are searched automatically. You search with one spelling and retrieve words with the same meaning spelled differently such as fiber and fibre. Both will be retrieved with either spelling. This is very helpful with databases from around the world, the vagaries of American and English spelling and is of assistance to those users whose first language is not English.

Compound Words - A compound word can be one word, hyphenated or two separate words. The distinction may differ from record to record, database to database or search to search. Now you can retrieve all three forms automatically no matter how you type in your search.

(c) Thesaurus

In the compilation of a database the keyword terms are controlled through the use of a thesaurus. The best keyword searching requires access to the thesaurus used by the indexers to allocate the preferred terms of their particular database. With the built in thesaurus the user is given magnificent access to new search tools in a completely interactive way. One can either inspect the thesaurus itself or use the facility to explode the search to cover the base term and narrower terms or related terms. It is even possible to determine the level using the tree display by counting the indented levels.

3. Choice of options

The software allows a great number of options which the user can set himself. One can set automatic plural searching or exact or varied word order in a phrase search. The display or output formats can be set by the user or the default settings can be used. The number of lines in the screen displays can be set from 25 lines to 43 or 50.

C. Anthology Publishing

NISC has been publishing compact disc databases since 1988. We concentrate on publishing databases in specialised fields often in areas with a small niche market. NISC have developed software to combine anthologies of smaller databases together and in this way make unique and valuable information resources available where the whole is far greater than any of the parts (5).

In such an anthology there will always be duplicate records, and NISC has developed duplicate record detection and composite record building software to deal with this. The unique parts of each record are retained and the resulting composite record contains the indexing perspective of each

data supplier. The aim of NISC is to present all the information in one place without the annoyance of repetition. Each source file is given credit for records in a composite record.

NISC has provided a means and a service to enable smaller database producers such as FISHLIT, to publish on CD-ROM and access the world market in a way which was not previously possible. NISC actively seek small specialised unique databases and is creative in producing combinations of information sources which delight users and database producers alike.

VII. INFORMATION FOR AFRICA

During the apartheid era, when South Africa was subject to academic boycotts and denied access to the information resources of the developed world, information professionals in this country filled the gap by producing information databases on many subjects. WATERLIT and FISHLIT are examples of these. Concurrently the database producers in the developed world turned their back on the research publications of this country and consequently many of the databases produced overseas do not offer information relevant to our South African needs nor do they cover the existing research in Africa well.

As Gibbs (3) says - "Many researchers in the developing world feel trapped in a vicious circle of neglect and - some say - prejudice by publishing barriers they claim doom good science to oblivion". Our mission at NISC SA is to help redress the neglect and ignorance of southern African and African research in mainstream science and humanities publications.

NISC SA, recently formed in June 1995, will publish four new CD-ROM titles this year, all of these include local African databases. These are discs with information for Africa about Africa and by Africans. In many cases these information resources are linked on the disc to useful databases from abroad.

A. Fish and Fisheries Worldwide

FISHLIT is now published on the CD-ROM "Fish and Fisheries Worldwide". This disc also includes 6 other databases from around the world. The South African database FISHLIT (1985-present) has 55,000 references and covers all aspects of fish and fisheries well, with unique material on Africa and useful references for developing countries. This disc includes information on aquaculture, fish diseases and parasites and fish nutrition in all 10,800 records on animal and human health matters.

B. Aquatic Biology, Aquaculture and Fisheries Resources

Using the technique of anthology publishing NISC has become the major publisher of databases in the fields of water resources, fisheries, oceans and aquatic studies. Recognising NISC's leading role in aquatic CD-ROM publishing the database producer Cambridge Scientific Abstracts have recently chosen to add their databases to the NISC stable. Two very important new discs have recently been published: "Aquatic Biology, Aquaculture and Fisheries Resources" (ABFR) and "Marine, Oceanographic and Freshwater Resources" (MOFR). These include the major database Aquatic Sciences and Fisheries Abstracts (ASFA) in combination with other relevant aquatic databases licensed to NISC including the South African database FISHLIT and aquaculture and fisheries subsets of Agris and CAB. These major collections together on two CD-ROMs provide unparalleled access to information on the science and management of aquatic organisms and environments both in the wild and under culture conditions. ABFR has all more than 35 800 health related references.

C. Water Resources Worldwide

In similar fashion NISC SA now publishes the disc "Water Resources Worldwide". The well known South African database WATERLIT is combined with SQUAREF produced by the Inland Waters Directorate

of Environment Canada and DELFT HYDRO from the Delft Hydraulics Laboratory, Netherlands. All these databases provide information on freshwater resources and their combination adds value to each of the other databases and offer more than 388,000 citations and abstracts on all areas of water research. WATERLIT (1975- present) has more than 237,000 citations and covers industrial and environmental aspects of water, wastewater and sanitation. Coverage of water in arid lands is particularly strong. SQUAREF (1970-1992) focuses on North American water resources and provides 86,200 references. DELFT HYDRO (1977-87) has an engineering and technological slant to the reference collection and contributes 64,600 records. There are in all more than 17,000 references on health related matters on this disc.

D. South African Studies

The new anthology "South African Studies" provides a unique combination of databases which give access to the documents published from South Africa. The disc includes the Index to South African Periodicals (ISAP) which is produced by the State Library in Pretoria, South Africa. ISAP has more than 145 000 records dating from 1987 to the present. More than 440 South African serial titles are indexed and many have abstracts.

Another State Library file on this disc is the South African National Bibliography (SANB). SANB has more than 44 000 records dating from 1988 to the present. This database provides a comprehensive record of South African publishing based on the material delivered to the State Library in terms of the Legal Deposit of Publications Act.

This disc also includes the Database of Theses and dissertations compiled by the University of Potchefstroom.

This CD-ROM covers a variety of topics including human and animal health and agriculture. A simple search reveals that

there are more than 1700 articles of relevance to the animal health practitioner. Specific searches will retrieve more articles. There are in all over 6500 articles on medical matters in general.

E. African Studies

The disc "African Studies", combines information from Africa, Europe and the USA. The multidisciplinary disc is set to become the premier information resource for those with an interest in the African continent. This disc covers political, socio-economic, international and development issues in the main. Included on the disc are the databases of the Africana Library of the University of California in Los Angeles (UCLA) (20 000 records), the Africa Institute in Pretoria, South Africa (60 000 records), the Sadius Database of the Institute of International Affairs in Johannesburg (50 000 records) and the Africana section from the library of the School of Oriental and African studies in London as well as the music database of the International Library of African Music (ILAM). We expect the subject scope and size of this disc to grow as new databases are added. There are in all about 4000 records of relevance to those with an interest in health matters, many of these cover traditional healing and community health matters.

Through NISC small and valuable databases are given a platform where they are able to take their place among the larger well known databases of the world. A database mounted on a NISC disc will be provided with the ROMWright search and retrieval software which is versatile, fast, powerful and user friendly.

VIII. "A COMPANY IN THE PUBLIC INTEREST."

NISC is a private company dedicated to serving the public interest. The company undertakes the publication of significant collections of information, for their popular research value and for dissemination in the scholarly and professional

community. NISC shares the library community's devotion to the advancement of knowledge worldwide and intend to contribute in four ways:

A. By encouraging the constructive use of information technology as NISC believes that access to information is essential for a peaceful world community.

B. Through the promotion of information of a social and environmental nature NISC hopes to promote sound ecological and social practices as awareness of global responsibility increases.

C. By enabling access to information which NISC feels to be vital to the public interest NISC offers some databases free of charge. The POPLINE database on population and community health is made available to developing countries free on request; the database on Child Abuse and Neglect is made freely available on request; and the ERIC database on education is given away free with the purchase of NISC discs.

D. NISC makes every effort to keep the cost of CD-ROM databases as reasonable as possible and does not believe in adding hidden costs or spending the customers' money on expensive brochures and up market premises.

E. NISC have seen the need to address the information needs of developing countries and have a programme to set up NISCs throughout the developing world.

IX. THE ROLE OF NISC IN AFRICA

NISC has recently opened a company in South Africa in order to serve the information needs of this continent. Africa is still the dark continent where information of relevance is concerned. African database producers should contact NISC to help the company to fill the very real need for information for Africa so that by the next decade we can demonstrate significant progress towards filling the void of information for Africa about Africa and by Africans.

REFERENCES

1. CRAMPTON, M. (1989) FISHLIT: A bibliographic information system for ichthyologists, fisheries scientists and fish culturists. In: Marine science information throughout the world. C.P Winn, R.W Burkhart and J.C. Burkhart (Eds). International Association of Marine Science Libraries and Information Centers 14th Annual Conference October 3-7, 1988: 91-99.
2. CRAMPTON, M. (1995) Publishing your database on CD-ROM for profit: the FISHLIT and NISC experience. *The Electronic Library* 13 (5):1-6
3. GIBBS, W.W. (1995) Lost science in the Third World. *Scientific American* August:92-99
4. JACSO, P. (1995) CD-ROM Cheers and jeers for 1995. *Information Today* December
5. NICHOLLS, P. (1995) A Private company with a public purpose: An interview with NISC's Fred Dürr. *CD-ROM Librarian* April :66-69