Networking databases: a user's perspective

Background

Computers have made a very significant impact on libraries and the work of librarians as they have on so many areas of science and technology. Communication networks have been developed to enable: (1) terminals to work with computers and (2) computers to link to computers. In the early days of computers they did just that, i.e. computed and little else. It was to ease the calculation load in such subjects as meteorology, nuclear science and general engineering that they were originally "invented". However, *commercial* users, e.g. banks, insurance companies, airlines, etc., quickly recognized their potential and applications were developed which used their capabilities. Commercial applications are very different from pure calculation: large amounts of information must be accessed and often searched but very little *calculation* is performed on it. The term "database" was introduced in this context.

Before the massive expansion of communications networks, each computer served those who occupied the same building and they often went to it and used it themselves. Then, in the 1960's, terminals were invented and these were located in offices in the computer building and hardwired to the computer. The user did not have to be in the room with the computer. The next stage was to locate terminals miles away from the computer they served. Modems, multiplexers and the like were incorporated and people could then use a computer in a remote town but permanent links were still required. Finally, flexibility was introduced and terminals could be switched to access a variety of different machines via communication *networks*. At this stage the hardware was in place and

G.N. Lance
Department of Computer Science
University of Bristol

applications software could be developed which utilized these facilities.

Databases

The definition of a database used here is "a collection of information acquired and used for a specific purpose". DBs may be numeric only, e.g. rainfall records for Edinburgh from 1900-1992 but, more usually, they are alpha numeric, i.e. they comprise letters and characters: ?, — etc., as well as numbers. Even before the *term* was invented databases existed, e.g. the hardcopy records of the Registrar General form a DB. Such records were stored on paper in filing cabinets, etc. With the coming of computers electronic means of storing DBs were discovered and there were, in sequence; paper tape, punched cards, magnetic tape, magnetic drums, cassettes and there are more recently, large magnetic discs or smaller CD-ROMS available to store our DBs. All such storage media were accessible by the computers to which they were attached and, hence, the terminals which could access the computers could also access the DBs. Since the terminals are networked, we can say the DBs are also.

DBs of interest to users of libraries

The subject must now be narrowed to address DBs from the point of view of the audience of librarians and publishers. Perhaps I should confess, at this stage, to being a very "computer literate" user and

hence I am not typical of all users of networked DBs. Furthermore, I invariably use academic — as opposed to public — libraries so my remarks may not apply to public libraries. The problems of non-academic libraries are very considerable because few, if any, can use JANET(1). My perspective of a librarian has changed during the 45 years or so of my working life. Originally, the librarian was a "keeper of books", the term included textbooks, novels, serials, etc. Today, this has all changed and most self-respecting librarians are now more accurately described as "information scientists". There are several reasons for this change: (1) the information explosion; (2) a recognition that users must be allowed and encouraged to exploit the valuable resources housed in libraries; and (3) the availability of computers, networks, terminals and DBs to assist the librarian with his or her task. The old-fashioned librarian had a reputation as a hoarder, i.e. he/she was not content unless the shelves were full of books: readers were allowed to borrow material rather reluctantly. Again, times have changed and I am pleased to say that the attitude is now much more along the lines that "the information I hold is no good unless the user can: (a) know of its existence and (b) access it". It is in the context of this change of attitude that DBs have been so useful.

There are two types of DBs used by libraries. The first provides the tools to "manage" the library, e.g. a list of users and their holdings, lists of suppliers of books, serials, etc. Some functions of this "management" software and the related DBs are relevant to users and are quite rightly made available to them, e.g. a user can access the system to determine which books he has on loan, which are overdue and he can place inter-library loan requests via the system.

The second type of DB is the one which is most used by readers. It comprises two main parts:

1.   Catalogue of material in the libraries.

2.   Bibliographic information.

These could be called briefly the "where" and the "what" DBs. In practice the student will consult the "what" first then the "where". In general a reader will be most interested in the holdings of his *local* library and this DB should be easily accessed by him. However, he might also wish to broaden the search for the work of interest and networks enable him to search other holdings in other libraries.

Until relatively recently, all these DBs were "hardcopy", i.e. in file indices or in books but now these types of DBs are stored electronically and the situation has changed radically. Not only can the librarian access this vast amount of information more easily but *so can the readers.*

The rôle of today's Librarian/Information Scientist

The discussion will now be narrowed still more to bibliographic DBs but the remarks also apply to citation indicies which are really a subset of bibliographic DBs. DBs in this category are typified by, for example, Medline in the U.S.A. or Inspec and BIDS here in the U.K. Some DBs are accessible via host computers e.g. Dialog, in the U.S.A., which provides access to many bibliographic DBs. There are now hundreds, some are accessible on only one host whereas others are available on many hosts. Basically, these list the contents of serials by author, subject, journal, date, etc. For them to be used by a researcher (that's me — a typical user!), it is necessary to pose the question in a suitable way in order to make a profitable search: it is essential to make some hits but you do not want to be drowned with results. Many librarians (information scientists) have specialised in producing such profiles for their users. Let us never forget that access to most DBs, other than those on our own computer, costs money and to get results back to your terminal also costs money so we don't want to flood ourselves with "junk lists".

It is fair to say that, if it were not for the

cost of access, librarians would have been overwhelmed with making search profiles for users. Even though this task requires more skill than just shelving books, it, too, can get tedious and it seems to me to be in the interests of librarians to allow and even encourage users to "do it themselves". Unfortunately, this can become prohibitively expensive for the uninitiated user. It is true that the *occasional* searcher will obtain more satisfactory results with the help of a librarian experienced in search techniques but I question whether librarians can provide such personalised help to the thousands of people doing bibliographic searches.

Let us now address the issue of citation indices, these were an "invention" of incalculable value to the researcher. It began as a manual exercise and hardcopy books of citations were published and sold. They are compiled from: (a) bibliographic records and (b) the references made to other papers in the listed documents. Today, all the necessary data are input to a computer and these can easily sort the data and produce the citation indices which are themselves a form of bibliographic DB.

Things have gone still further, in the recent past, ISI have made available *machine readable* versions of their bibliographic and citation DBs. In this country, this is being made available by Bath University in a service called BIDS (2). BIDS is accessible to members of the academic community who also have use of a terminal which is part of JANET — the academic communications network. Each Institution which accesses BIDS pays an annual fee for the privilege and this helps to cover: (a) the cost of the DB per se and (b) the cost to Bath University of running the service. However, the important point is that the charge is *not* proportional to use — but this could always change! Although BIDS has been available only one year it has generated considerable interest and its use is increasing very rapidly (3).

With the financial inhibition removed, an objection to the end user doing searches for himself is also removed and most librarians are now faced with the problem of training their customers how to use BIDS for themselves. In the long run this may be doing themselves out of the job of profile construction but their real rôle as information scientists is not affected.

Training of users is in itself a major problem. Bath University has produced a "BIDS Instruction Pack" (4) which is intended for those faced with user training. In addition thousands of "User Guides" have been printed which summarise the procedures. However, the best approach is to incorporate adequate "help" information within the system itself and in BIDS this approach has been adopted so that "help" can be invoked at any stage of the search procedure, if the user does not understand. Unfortunately, the help system is not ideal for teaching the new user the ideas and principles behind DB searching and this task remains one for the librarian.

The rôle of the end user

At last I can address the full title of this paper! I truly believe that access to BIDS, and systems like it, by the end user will have a major impact on research. It is now so easy to do literature searches by topic or author that any excuse for reinventing the wheel has gone for ever. There is a danger in this ease of access because the uninitiated may well be unaware of pitfalls which lurk. Has he allowed for synonyms? Will the use of the wild characters "*" in e.g. prog* produce too many hits?

More importantly, if the search has been done properly it will become clear what work has *not* been done and so ideas for research can be generated. However, and this is the crux of my thesis, this ability is of most value to the researcher himself. Provided always that he can produce accurate information by searching himself then he will not need to call on the services of his local friendly information scientist. Thus everything hinges on whether the user's search is as accurate as one obtained

using an intermediary. My belief is that this should be possible provided that he has been able to learn the techniques. I am convinced that it is in the interest of librarians to help with such training.

It goes without saying that BIDS must be easy to use because using it is *not* the full-time occupation of the typical researcher. He will forget most of what he learned about it in the interval between use. Hence, the menu and help facilities provided are essential if the end user is to make effective use of the system. But rather more is needed, the user should understand the , nature of the DB and how to construct a search which will give an accurate and full response to his perceived question. In this regard he relies heavily on the librarian — or information scientist — because BIDS provides no "help" in this particular regard.

The earlier that people become acquainted with bibliographic DBs, citation indices, etc., the better. Certainly, undergraduates should be taught the services available and how to take advantage of them. It is in this context that librarians will come into their own, they should be the trainers. Fortunately, BIDS has recognized this and, with the help of librarians (and a few users) useful training material has been made available.

The future

Some time has been spent on describing how we got where we are today and the present state of play; so I will conclude with some star gazing, some prophesies and provide a wish list.

Let us assume that as a user of BIDS, I find some references which are relevant to my interests. So what? All I have is titles, authors, place of publication and dates. If I am very lucky, I might have an abstract too — these are available in e.g. The Computer Literature Database-CLDB (5) or from abstract journals. If I wish to pursue my enquiries, then I need to see the paper itself and the problem has only just begun.

I access my *local* on-line catalogue, i.e. Libertas, no joy! I then exit from Libertas and access, via JANET, other possible locations and perhaps find that the paper is in the London University Senate House Library. (Barfoursh et al (6) have done some interesting work on accessing *all* libraries available from JANET.) Back to Libertas to request my *local* library to obtain the Journal via Inter-Library Loan. I understand that, before long, it will be possible for users to invoke JANET from within Libertas. Clearly, this will be of considerable benefit to the typical user who would not otherwise have ready access to JANET. However, there is a very nasty sting in the tail here! Have you ever contemplated what an increase in ILLs will be generated by this facility?

The ideal would be to be able to access, on-line, the full paper required. For this to be possible it needs to be "published" electronically. As an example, I see it working as follows: The Cambridge University Press publish the *Computer Journal* and *they* store all the papers in the *CJ* on *their* DB. I could then access this DB to read (or acquire a copy of) the paper I need. Naturally, CUP would make a charge for this service, and so they should, because such a service may well, in the long-term, completely replace hard copy versions. The ADONIS project supported by Elsevier, Pergamon, Blackwells and others has gone some way to addressing this issue but the data are provided on CD-ROMs.

The implication of this suggestion is that *all* publishers will become holders of DBs, containing their publications, which can be accessed by users via a public communications network. Whether this vision ever becomes a reality is debatable. If only a few publishers co-operate, the users will not find it worthwhile to equip themselves with the necessary hardware but, once a "critical mass" of publishers is reached, the remainder will find they *must* co-operate in order to compete. Who will volunteer to be the first to begin this

exciting new phase of "publication"? Perhaps Elsevier will go one step further?

Electronic publishing raises big questions regarding copyright and the infringement thereof, we have seen this already in the case of photocopiers. How will charging be implemented? Any charging method adopted must be such that it

(a) Allows the publishers to provide a viable alternative to the printed versions of publications and still remain profitable and

(b) Does *not* make it impossibly expensive for teachers, researchers and undergraduates to make use of on-line publications.

It is to be hoped that librarians and members of UKSG will see the advantages for them in this new situation. No more hardcopy journals to find shelf space for, no more expenditure on journals which seldom get read, no more inter-library loans (remember this idea could extend to books as well to periodicals). In fact, if the terminals are in users' own offices, or public areas, they need never even visit the Library. Utopia?

An objective of the UKSG is to ". . . assist in the developement of appropriate research in the field of serials management". May I respectfully suggest that this idea, which is undoubtedly controversial and provocative, is so important that it is worthy of further study.

The views of all concerned namely, publishers, librarians and users must be taken into account because all three of these groups must be happy with any scheme which is developed, if it is to succeed.

References

1. "Workshop of Networking and Public Libraries" in *JUGL Newsletter*, No 5, Winter 91/92

2. Scanlon S., BIDS: The Revolution in Database Access, *Serials*, 5, No. 1, pp18-28, March 1992

3. A report of the JUGL/UKOLUG Joint Meeting on "Using JANET and the Bath ISI Data Service" *JUGL Newsletter*, No 5, Winter 91/92

4. *BIDS Instruction Pack*, Bath University Computing Services, Bath 1991

5. *CLDB — The Computer Literature Database, User Guide*. Dept. of Computer Science, Birkbeck College, London WC1E 7HX

6. Barfoursh A.A., Lance G.N. and Rogers M.H., Multi DataBase Enquiry Service MDB-ES, *Proc. of the Data Base Conf.*, Melbourne, Feb. 1990