

Information on demand: online retrieval from external databases

J. A. Large

College of Librarianship Wales, Aberystwyth, UK

ONLINE INFORMATION RETRIEVAL

With just one telephone call it is possible to gain access to vast stores of information on computers which may be located anywhere in the world and to find the item of sought information within a few minutes. All that is needed, in addition to a telephone, is a terminal of some kind and, as Barry Mahon explained in the previous paper, a device called a modem which connects the terminal to the telephone line. Usually the terminal will be a microcomputer which has an external port to enable it to be connected with the modem. A communications software package will also be needed in order to communicate with other computers. An investment of between £1,000 and £1,500 (including a microcomputer and printer which can be used for other purposes) provides the user with access to a variety and quantity of information only to be found in the very largest libraries.

Online information can be valuable to all kinds of users. It is widely exploited in libraries and information units, of course, but other professionals such as business and financial consultants increasingly have recourse to online data. Translators must often be faced with the need to find information quickly and that information might be extremely varied in content. Standard reference books may be on hand, but are they up-to-date and can they always satisfactorily answer the information request? As one technical translator has noted:

A translator of scientific-technical material must be a very versatile animal if he is to survive, perhaps today more than ever before, when some technologies are running five to ten years ahead of their lexicographers.

Translation is a vital part of the dissemination of information, and translators are finding that they must learn the skills of information scientists, not only to be able to locate information on both the mundane and the arcane, but also to find equivalencies on the other side of the language barrier.¹

Online information systems, which are an everyday part of the information scientist's world, also offer considerable scope to the information-seeking translator.

DATABASE VENDORS AND SERVICES

Online information is organised in databases, of which there are several thousand available to anyone who acquires the necessary equipment and signs a contract to obtain the necessary passwords. In some cases the database producer offers the database directly to users (as is the case with the *Wall Street Journal*, published by and available online from Dow Jones). In most cases, however, the database is offered to the public by an intermediary organisation variously called a vendor, host or search service. The vendor provides the computing facilities and customer support services through which various databases from different producers can be used. Some vendors are described as online supermarkets and provide large numbers of databases: the Californian vendor, Dialog Retrieval Services, for example, offers around 300 databases to its customers. Table 1 gives some idea of the growth of online databases and vendors over the past eight years. Others offering a wide portfolio of databases include the US vendor, Bibliographic Retrieval Services (BRS); the UK vendors, BLAISE-LINE and Pergamon Financial Data Services; the European Space Agency's Information Retrieval Service (ESA-IRS); ECHO, offered by the Commission of the European Communities; the Swiss vendor, Data-Star; the German host, Dimdi; Data Centralen from Denmark; and the French vendor, Télé-systèmes-Questel.

Year	Databases	Vendors
1979-80	400	59
1980-81	600	93
1981-82	965	170
1982-83	1,350	213
1983-84	1,878	272
1984-85	2,453	362
1986	2,901	454
1987	3,487	547

Table 1. Growth of online databases and vendors

DATABASE TYPES

Online databases are often divided into the following categories depending on the type of information they contain: reference databases which are either bibliographic or referral and source databases which are either numeric or full-text.

Bibliographic databases contain information about publications such as books, articles, reports, theses, conference papers, patents and audio-visual productions. By mid-1986, according to Hall, 159 million references were available in bibliographic databases with an annual increase of 16 million. Although there is considerable overlap between the various bibliographic databases, Hall estimates that 100 million of these references are unique. These figures should be compared with his estimate that even the largest libraries in the world probably hold information on only 10 to 20 million documentary units.³

All the major disciplines are covered by bibliographic databases. Some of these are very large; Biosis Previews, for example, has around five million references to literature on the life sciences and each month 40,000 more are added. Of particular interest to translators are likely to be three much smaller databases: World Transindex (on ESA-IRS), produced by the International Translations Centre and CNRS and containing 250,000 references to translations; Sigle (on BLAISE-LINE), a database of references to grey literature, including translations of scientific and technical reports and articles held at centres in Belgium, Ireland, Luxembourg, Sweden, West Germany and the United Kingdom (British Library Document Supply Centre); and Linguistics and Language Behaviour Abstracts (on Dialog and BRS) which has around 90,000 references.

Referral databases are the online equivalents of printed reference works such as directories, dictionaries and encyclopaedias. The relative ease of updating entries mainly accounts for the large number of directories found online. Agrep, for example, lists agricultural research projects in EC member countries and is updated monthly; Business Software Database describes 5,000 business systems and utility software programs for all types of computer.

Dictionaries are less common online, but in the chemical field there are a number of chemical databases which bring together the various synonyms for the same chemical. Translators may be more interested in a database such as Eurodicautom, the multilingual terminology database (available from the Commission of the European Communities on its host, ECHO) which contains 500,000 terms and phrases and their translation equivalents in one or more Community languages, together with more than 120,000 abbreviations.

A small number of encyclopaedias are now available online, including *Everyman's Encyclopedia* and the *Academic American Encyclopedia*. The

absence of illustrative material, of course, is a considerable drawback, and in practice these databases seem to be updated infrequently. Nevertheless, it may be possible to find information online which would prove elusive in the printed equivalent (see below).

Source databases, both numeric and full-text, are of growing importance. There are very many *numeric* or textual-numeric databases which provide demographic information (eg CENDATA with data from the US Census Bureau) or business information (eg Jordanwatch's data on individual companies). Interesting examples of the *full-text* variety are to be found in the current affairs field; the contents of newspapers such as *The Guardian*, *Financial Times* and *Daily Telegraph* from the United Kingdom or the *Boston Globe* and the *Philadelphia Enquirer* from the United States can be searched online and the full text retrieved. Generally they are updated daily but are one day behind the printed version. A growing number of wire services such as Reuters, Tass and DPA (Deutsche Presse-Agentur) are also available online and in many cases are updated every 15 minutes. Some journals, such as the *Harvard Business Review* and *The Economist* are also available online. The retrieval capabilities of online systems means that these databases can now be used as rich sources of retrospective information as well as sources of current information. The printed newspaper can be browsed on the train to work, but the online version can be searched to find a wealth of information buried in old issues which could never successfully be uncovered even in the case of those relatively few newspapers which publish printed indexes.

SELECTING THE RIGHT DATABASE

With so many databases available, it is by no means straightforward to identify the best database for any particular information need. Directories such as those by Williams⁴ and Cuadra⁵ are useful, and both these examples are available online as well as in print. Advice on database selection as well as guidance on getting the most from those databases can be found in a manual edited by Armstrong and Large⁶. Some vendors such as Dialog and ESA-IRS provide online index files to their databases. Familiarity with the printed versions will also help in selecting the best database, though it should be noted that the printed and online versions are not always precisely equivalent. In some cases the information is updated at different periods and intervals and the actual content may also vary; Psychinfo, for example, is the online version of *Psychological Abstracts* but online coverage is broader than in the printed form.

ADVANTAGES OF ONLINE SYSTEMS

In some cases information is only available online and there is no print version: one example is the management database, ABI/INFORM. Where the two media are available the user will want to evaluate the advantages and disadvantages of online information sources as compared to the print versions and make a choice.

The first thing to notice is the speed at which online information can be found. A typical search on a bibliographic database such as CA Search, which contains more than eight million chemical references stored on a computer located on the other side of the world, will normally take about 10 minutes. This must be compared with a laborious manual search through volume after volume of the printed equivalent, *Chemical Abstracts*.

Secondly, online databases typically provide many more access points by which information can be traced than would the printed work. Often it is possible to search for words or phrases in a title, an abstract or in the body of the text (in the case of full-text databases). Thus searching is not limited to index entries selected by the indexer. It is also often possible to search on other features such as the language in which an item is written, and its date or place of publication. To take one example, the printed *Everyman's Encyclopedia* is arranged alphabetically by subject headings. Using these headings it is a straightforward matter to find entries for the two Russian poets, Lermontov and Pushkin, both of whom had the misfortune to be killed in duels. It is a very different matter, however, to find information about any Russian poets killed in duels: this can only be found if the poets' names are already known. Short of recalling the names of as many Russian poets as possible and checking each entry for causes of death, the question cannot be answered. Online, such a request can be answered in a couple of minutes by isolating those entries which contain the words *duel*, *duels*, or *duelling*, and *poet*, or *poets*, and *Russia*.

Online information can be brought up-to-date relatively easily as compared with print, and some databases are updated every few minutes, although more typically online updating will occur monthly or quarterly. However, not all online sources are updated more frequently than their print equivalent. It can be safely assumed that a database like the Bible (King James' version) will never be updated.

Another feature of online is that charges are normally only incurred when a database is used: online costs are directly related to searching*. The precise method of charging varies according to vendor and database, but is generally related to the length of time spent online and the amount of information retrieved. Thus, a database which is infrequently used will

* Many vendors now only charge for records visualised and/or printed – and not for time searching. This further reduces cost.

cost very little, but of course a heavily used database may generate large bills. This is in contrast with printed works whose initial purchase price is unrelated to subsequent usage but which may require subsequent expenditure for purchasing revised and updated editions.

Retrieved data can be manipulated in various ways. Apart from obtaining a printout of search results which obviates the need to take notes, searches can be downloaded from the vendor's computer onto the searcher's local computer. The downloaded data can then be edited, using word processing software, in order to delete unwanted data or to add annotations. For example, a list of downloaded references can be converted into an annotated bibliography. With the right kind of local software it is also possible to transfer downloaded information into an in-house information retrieval system. It should be noted, however, that downloading is often subject to contractual conditions as established by the vendor and database producer⁷.

DISADVANTAGES

The advantages of online sources must be balanced against a number of disadvantages. Online search costs are typically quite high and very visible. They comprise three elements: the telephone call to the nearest data network node, which for many users will be a local rate call; the data network rate, which is largely related to time but also includes an element for the volume of data transferred in each direction (but not for distance); and the charge made by the vendor, which varies according to both vendor and database. A search on a relatively cheap database is likely to cost around £1 per minute, with a typical search taking 10 to 15 minutes, but this rate can be much higher. Online costs are probably the greatest barrier to more extensive usage.

There is an ever-present threat of technical problems preventing or interrupting a search. Local equipment failure should be uncommon, but problems may arise from time to time with the vendor's equipment. The most fragile element, however, is the telecommunications link, as Barry Mahon has also explained. In 1987, Eusidic (European Association of Information Services) monitored over 5,000 calls made during one week and found a failure rate of almost 30 per cent⁸. It should also be noted that vendors are not available at all times. For example, BLAISE-LINE only operates during normal office hours, five days a week, and most vendors close down for public holidays. Such restrictions particularly affect searchers who are self-employed and who do not work conventional hours.

A third problem relates to the methodology for using online services. At present, a proportion of the search cost is directly linked to the amount

of time spent online. This favours users who can formulate efficient search strategies and requires a certain amount of skill. Furthermore, for many online systems queries have to be formulated in a command language which has its own vocabulary and syntax. Although command languages are very limited in scope and therefore not too difficult to learn, there are a lot of them as each vendor tends to use its own. Some vendors have eliminated the need to learn a command language by creating menu systems: users issue instructions by choosing options from menu screens. However, such menu-driven retrieval systems tend to be slow and rather tedious to operate.

TRAINING

The would-be online searcher must learn the rudiments of retrieval strategies and search procedures if the full benefits of online systems are to be reaped. If the vendor does not provide menus, users must master the command languages. It may be possible to learn from colleagues, or alternatively to attend a training course. Vendors generally offer regular training courses, often for one day, at both introductory and advanced levels. Some of the larger database producers also provide courses. Professional bodies such as Aslib also run regular training seminars and the UK Online User Group arranges seminars on various aspects of online searching. The United Kingdom also has many local online user groups which run seminars and offer opportunities to meet other people locally involved in online searching.

Some vendors such as STN and Data-Star now provide training software which can be used on the client's own microcomputer; others, such as Dialog, provide online training files which are very cheap to use. A number of microcomputer software packages are now available which are designed to help the user construct their search strategies, and thereby reduce the need for training. Alternatively, it is possible to establish a telecommunications link with an intermediary gateway service such as EasyNet which assists in strategy construction and database selection before automatically connecting the user to a suitable vendor⁹.

There is also a wide range of published material to assist the searcher—textbooks, journals (such as *Database*, *Online*, and *Online review*) and handbooks such as the useful guide to the command languages used by the major hosts¹⁰. Aslib regularly publishes an invaluable guide called *Going Online* which contains a wealth of information on all aspects of online searching¹¹.

VIEWDATA

Viewdata systems, of which Prestel is the UK example, provide an online link between the user and data stored on a remote computer. While they differ from the online systems described above, in terms of their historical development, information content and retrieval mechanisms, these differences are beginning to diminish.

Viewdata was developed in the late 1970s by the then British Post Office to boost home use of the telephone network using two pieces of equipment which were by then common in homes – the telephone and the television. Users required a modified television set and a numerical keypad which was used to select options from on-screen menus until the sought information was found. Charges were made for some, but not all of the information screens, and of course telephone calls were charged at the usual rates.

The information provided on Prestel was intended to appeal to the home user, but little thought had been given to their requirements. The cost of the modified television sets and the telephone calls combined with the relatively unappealing nature of the information account for Prestel's very limited success in its early years. More recently, however, viewdata systems have shown signs of healthier growth. A number of countries such as France and Canada have successful services while Prestel itself seems to be making progress albeit more in the business than in the home sector. Microcomputers with a viewdata card have replaced the television set and a keyboard has replaced the old keypad.

Information retrieval is better organised and more sophisticated and the information available is now of a more serious nature. For example, it now provides financial data though this may only be available to closed user groups who pay an additional premium. Prestel can now be used as an electronic mail service¹². Local viewdata services operated by local authorities are quite common and offer online access to community, travel and business information.

CD-ROM

CD-ROM (Compact Disc - Read Only Memory) is an in-house retrieval technology which does not rely upon telecommunication links with an external computer. In this respect it differs markedly from the online systems, including viewdata, described above. In other respects, however, it does share common characteristics. In both cases the data is normally provided by an external agency – a database producer – rather than being generated internally by the client (although it would be possible, albeit expensive, to commission a CD-ROM which could store local information such as a library catalogue). Several databases exist in

online and CD-ROM form and the information retrieval methods are similar. Indeed, some online vendors such as Dialog also market CD-ROM products and supply the search software for them¹³.

A CD-ROM looks just like an audio compact disc – it has identical dimensions and is manufactured in the same way. Information is stored in pits on the disc and is read by a low power laser. Its storage capacity is very impressive – around 550Mb (which can hold the equivalent of 250,000 A4 pages or the contents of about 1,500 floppy discs). This is sufficient to store the entire contents of a medium sized database on one disc. The only drawback is that the disc can only be read; it is not possible to erase information from or add information to the disc. The laser player costs around £600. This is linked to the microcomputer which is equipped with a special card costing around £600 and uses search software supplied with the CD-ROM. Searching is usually carried out via menus rather than a command language and search times are a little slower than for online.

One very significant feature of CD-ROM is that payment is not related to usage; a payment or subscription is placed for the product as in the case of printed material. Some CD-ROM products cost no more than a few hundred pounds, but some carry a very high price tag, in excess of £1,000. Only relatively heavy use could justify their purchase.

Around 300 databases are currently available on CD-ROM although this figure is growing quite rapidly. The databases fall into much the same categories as online databases, and include bibliographies (*Medline*), dictionaries (*Oxford English Dictionary*) and directories (*The Postcode Address File*). Of particular interest to the translator are likely to be *Harrap's Multilingual Dictionary*, which contains 13 dictionaries from six publishers in a variety of languages including Chinese, Dutch, English, French, German, Italian, Japanese and Spanish, and the *International Dictionary of Medicine and Biology* from Wiley¹⁴.

Currently, CD-ROM and online databases are seen to be in competition with each other as well as with print; the future will show whether they can co-exist.

ONLINE AND THE TRANSLATOR

The variety of information online probably means that most translators will find something of relevance. The databases which specifically deal with translation activities are clearly of central importance; other databases also offer an opportunity to check on word usage, to find articles by particular authors, or to check facts. The costs may seem high because they are immediate and direct, but against this must be set the saving in time and the immediate accessibility of the information. For the busy translator, online searching can play a valuable part in the daily working schedule.

REFERENCES

1. Crump, Ted. Scientific/technical translations in a research library. In: Katz, Bill and Fraley, Ruth (eds), *International aspects of reference and information services*. New York: Haworth Press, 1987, pp 113-122.
2. Cuadra, Carlos (ed.), *Directory of online databases*. New York: Cuadra/Elsevier, 1987.
3. Hall, James L. *Online bibliographic databases: a directory and sourcebook*. 4th ed. London: Aslib, 1986.
4. Williams, Martha E. *Computer-readable databases: a directory and data sourcebook*. Amsterdam: North-Holland, 1985.
5. Cuadra, *op. cit.*
6. Armstrong, C.J. and Large, J.A. (eds), *Manual of online search strategies*. Aldershot: Gower, 1988.
7. Tenopir, Carol. Online searching with a microcomputer: downloading issues. *Microcomputers for Information Management* 2(2) 1985, 77-89.
8. May, Nigel. European networks make a small improvement. *UK Online User Group Newsletter* 52, Nov-Dec 1987, 26-27.
9. O'Leary, Tim. EasyNet: doing it all for the end-user. *Online* 9(4) 1985, 106-113.
10. Arthur, Adrian (compiler), *Quick guide to online commands*. 2nd ed. London: UKOLUG, 1988.
11. Starkey, Helen and Thwaites, Beverley. *Going online 1989*. 7th ed. London: Aslib, 1988.
12. *Aslib Proceedings* 37(6/7) 1985, entire issue.
13. Hatvany, B. Comparison of CD-ROM and online. In: *Proceedings of the 11th International Online Meeting, London, December 1987*. Oxford: Learned Information, 1987, pp 285-290.
14. *The CD-ROM directory 1988: the international directory*. 2nd ed. London: TFPL, 1987.

USEFUL ADDRESSES

Aslib, The Association for Information Management, 20-24 Old Street, London EC1V 9AP

Eusidic (European Association of Information Services), 1st floor offices, 9/9a High Street, Calne, Wiltshire.

UKOLUG (UK Online User Group) c/o Secretary, UKOLUG, Information Automation Ltd, 'Penbryn', Bronant, Aberystwyth, Dyfed SY23 4TJ.

AUTHOR

Dr J. Andy Large, College of Librarianship Wales, Aberystwyth, Dyfed SY23 3AS.