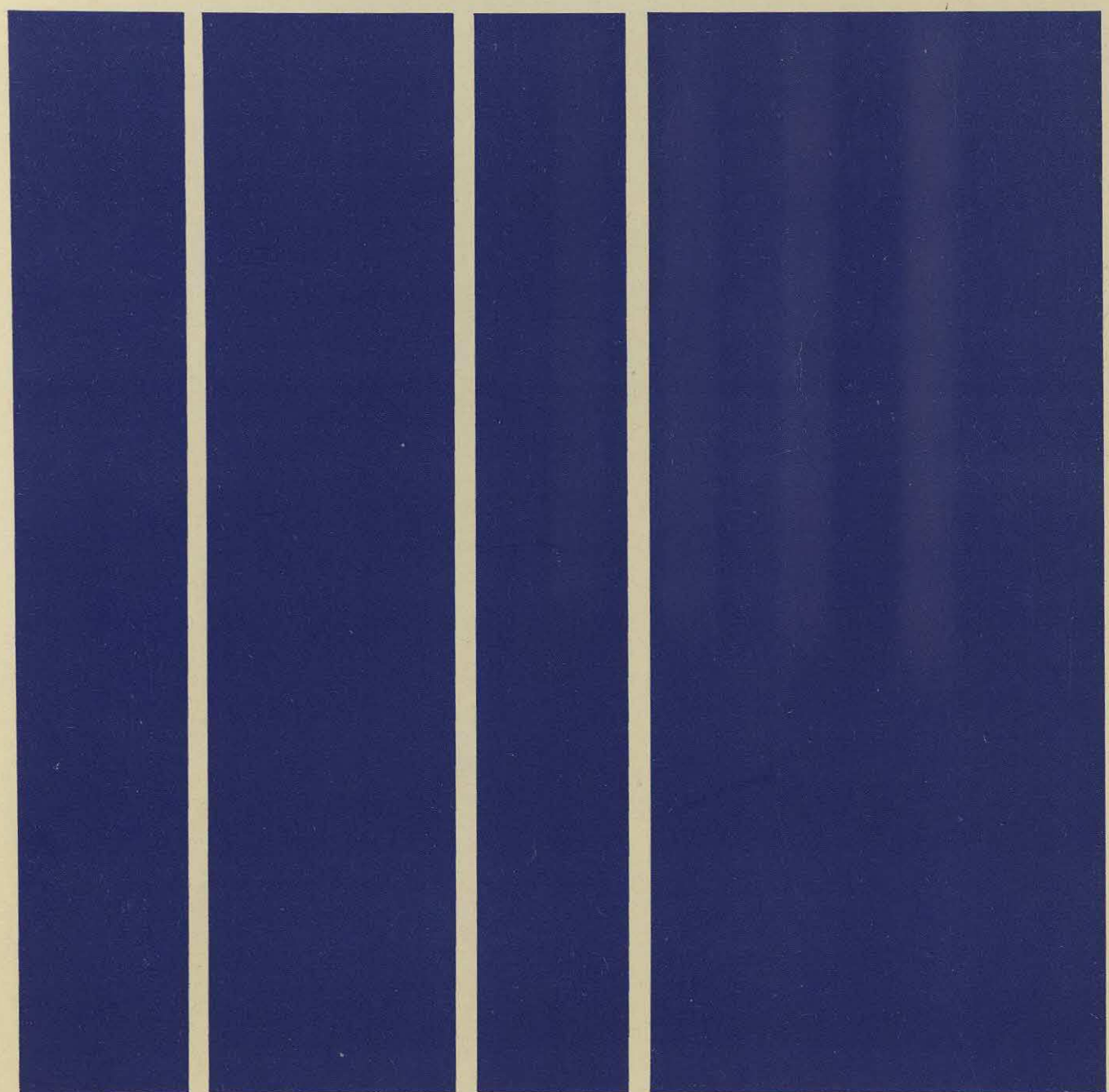
Report Series
No 10

Public On-line Information Retrieval Services

John Kinnear

November 1978



The Butler Cox Foundation

Abstract

Report Series
No 10

Public On-line Information Retrieval Services

by Peter Appleby
November 1978

ABSTRACT

On-line information services are now appearing in Europe, and these offer all organisations new ways to obtain information from outside. This external information is needed by managers and professional staff in almost all parts of an organisation.

These information services, which enable an end user to obtain information which is relevant to his needs, represent one way of overcoming the problem of an ever-increasing supply of published material.

The report examines the development of information services in the US, Europe and Japan and discusses the directions in which those services can be expected to expand in the next few years through a number of relevant technological advances.

The report offers advice to organisations on the action they need to take to ensure that they make full use of the currently-available services, and that they equip themselves to exploit new services as they reach the market-place.

The Butler Cox Foundation

The Butler Cox Foundation is a research group which examines major developments in its field — computers, telecommunications, and office automation — on behalf of subscribing members. It provides a set of 'eyes and ears' on the world for the systems departments of some of Europe's largest concerns.

The Foundation collects its information in Europe and the US, where it has offices through its associated company. It transmits its findings to members in three main ways:

- As regular *written reports,* giving detailed findings and substantiating evidence.

- Through *management conferences,* stressing the policy implications of the subjects studied for management services directors and their senior colleagues.

- Through *professional and technical seminars,* where the members' own specialist managers and technicians can meet with the Foundation research teams to review their findings in depth.

The Foundation is controlled by a Management Board upon which the members are represented. Its responsibilities include the selection of topics for research, and approval of the Foundation's annual report and accounts, showing how the subscribed research funds have been employed.

Report Series No. 10

PUBLIC ON-LINE INFORMATION RETRIEVAL SERVICES

by Peter Appleby

November 1978

Butler Cox & Partners Limited Morley House 26 Holborn Viaduct London EC1A 2BP

TABLE OF CONTENTS

I. THE SIGNIFICANCE AND POTENTIAL IMPACT OF ON-LINE INFORMATION RETRIEVAL SERVICES

The quality of almost all business decisions is highly dependent upon the information used in the decision making process. High quality information – by which one means relevant, timely, accurate and comprehensive information – is therefore crucial to the continued success of all organisations. One source of such information is published material available to all who require it.

Such information is crucial in such areas as product planning, product engineering, market research, legal departments and information systems planning. Questions such as 'what patents or products already exist?' and 'What are the qualities and the availability of the materials we could use?' can be difficult to answer well.

The 1973 oil crisis added an additional dimension to the situation, since the radical effect that it had on the world economy has introduced added risks to all major business decisions. The instability which the crisis caused (and which appears likely to be with us for some time) has resulted in information, and therefore the decisions based upon it, becoming outdated much sooner, and in unexpected ways, than managers were previously accustomed to expect.

On-line information retrieval services appear to offer in the 1980s one important way of compensating for such changes since, potentially, they offer the opportunity to retrieve high quality information considerably more economically than would otherwise be possible, if indeed it would be possible at all.

A The Likely Growth of Information Retrieval Services

In Europe few organisations have, so far, made very great use of on-line information retrieval services. Typically, large UK-based organisations use these services for no more than three hours a week, and taking one advanced user as an example – ICI – the weekly use is only about twenty-five hours.

In the US, by contrast, organisations started to use on-line information services earlier, and have made much greater use of them.

The pace of development and the usage of information retrieval services is now accelerating. In the US particularly, the number of organisations that are respectively producing, selling and using databases grows daily. In the UK the planned viewdata service, Prestel (which is described in the Section beginning on page 24), is a development that will create a much greater awareness of scope for, and opportunities with on-line information retrieval services. The European network EURONET (which is described on pages 15 to 17) will undoubtedly produce a dramatic increase in the use of on-line information retrieval services. Much of this report is concerned with the present state of the art in the US and Europe, and with the future of on-line information retrieval in those continents.

Substantial developments have also taken place in Japan, and in that country too there is likely to be a rapid increase in the use of on-line information retrieval services. Again, the present

position and likely developments in Japan are discussed later in this report.

Some of the new technologies that are on the horizon (such as video disc, personal microcomputers, parallel processing computers and satellite communications facilities, which are all discussed in the Section beginning on page 28) could all add to the speed of development and the growth of usage of on-line information retrieval systems.

B Potential Opportunities and Potential Threats of On-line Information Retrieval Services

The concept of storing information in a machine-readable form so that it can be accessed quickly and easily through a terminal device poses both opportunities and threats to many sectors of business.

The facility of being able to have access on-line to large, comprehensive information databases represents a new and powerful tool to managers, scientists and technologiests in business, as well as to many individuals in the professions, and in educational and research establishments. In a business organisation, however, there is one potential danger. Different individuals and departments in different areas of a company may learn about on-line information retrieval from different sources and in different ways. The potential danger is that the separate individuals and departments within an organisation may, through ignorance of, or disregard for, the needs of others, embark on setting up their own route to information retrieval systems. Unless the requirements of all the areas of a company are co-ordinated this approach will obviously lead to duplication and waste of effort.

As on-line information usage grows, it may reduce or eliminate the requirement for certain printed material such as periodicals. The economic considerations of switching from hard copy to on-line will become more important as the usage of on-line information retrieval services grows. From the viewpoint of an organisation that accesses and uses the information, the potential benefits of cost savings in areas such as periodical subscriptions, book purchasing, and so on could be substantial.

The ramifications of such savings could result in a related and additional effect. As people within an organisation become aware of the power of on-line information retrieval services, more and more sectors of the organisation might decide to use the facilities of on-line information retrieval systems. The tangible benefits of financial savings would then be supplemented by intangible benefits, such as a better informed organisation, a quicker response to information needs and so on.

The on-line information retrieval services could therefore represent much more than simply a better way of storing and accessing information. These new facilities could have a radical effect on the way in which organisations operate. Improved information services could stimulate both management and professional skills to a new, much higher level of efficiency and effectiveness.

Many publishing organisations and organisations with interests in the dissemination of information on paper, already realise that on-line information retrieval services may, in the future, produce a threat to their present business, and are now taking steps to protect their future interests by evaluating their situation in the light of the present and the likely developments.

The likelihood that more and more data will be held on computer files has raised many debates on privacy and security. These debates are particularly centred on the contents of the files and the question of which individuals should and should not be able to access the files.

At present, very few countries are taking enough action to control the storage of data on computer. So far as legal controls are concerned, Sweden has probably advanced further than any other country.

One aspect of privacy and security that is of concern to industry and commerce is that of industrial espionage, and the possibility that one company can study the nature and the pattern of a rival company's search. It can sometimes be as revealing to know what search patterns a company is following as it is to know the contents of the individual files that the company is searching.

II. THE SCOPE, PURPOSE AND INTENDED READERSHIP OF THIS REPORT

A The Scope

Having regard to the developments and issues discussed in the previous section, the following major questions require attention:

1. Is on-line information retrieval a subject that is sufficiently important to warrant organisations devoting attention to it now?

2. If so, what action should management services managers be taking, both now and in the longer term, to prepare for on-line information retrieval?

3. When is on-line information retrieval likely to have a significant impact on business organisations and on their day-to-day operations?

This report sets out to answer these questions by looking at the present state of the art in the US, Europe and Japan, by looking at both present and future developments in technology, and by looking ahead at the way in which on-line information retrieval services are likely to develop in Europe.

B The Purpose

This report has three main purposes.

Its first purpose is to assess the way in which on-line information retrieval services are expected to develop in Europe, based both on the world situation of on-line retrieval today in the US, Europe and Japan, and on expected developments in hardware and software.

Its second purpose is to highlight the major considerations for users and potential users of information retrieval services.

Its third purpose is to identify the role that the management services function in an organisation should play in order to ensure that a coherent, effective and economical information strategy is established within the organisation.

C Intended Readership

The report is intended to be read by all those who are responsible for ensuring that the supply of information within an organisation (and particularly the supply of information that is acquired from outside the organisation) is, and continues to be, adequate. It is intended particularly for those within the management services function of organisations, since it is they who are most usually consulted for advice on how to install and exploit an information system. It is also intended for librarians and others who have a responsibility for providing information.

III. BACKGROUND TO ON-LINE INFORMATION RETRIEVAL SERVICES

A The Need for On-Line Information Retrieval Services

Most large and many medium-sized companies in commerce and industry today have one or more information or intelligence departments or units. As part of their function these departments or units usually have one or more staff members who are engaged in trying to satisfy the information needs of the particular organisation by scanning books, professional journals, periodicals, newspapers, reports and other kinds of published information.

Over the past decade there has been a steady, if not spectacular, increase in the variety and volume of published information. For example, as Figure 1 illustrates, between 1971 and 1978 the number of published periodicals grew by 20%. The growth in the variety and the volume of published information has increased the burden that the task of scanning printed material represents. The burden has increased also because it becomes increasingly difficult in any organisation to classify and index the printed material in such a way as to best satisfy the diverse and yet inter-related needs of the different functions and areas of interest in that organisation.

GROWTH IN THE NUMBER OF PERIODICALS

YEAR	NUMBER OF PUBLICATIONS ACCORDING TO ULRICH'S INTERNATIONAL PERIODICALS DIRECTORY
1971	50,000
1973	55,000
1975	57,000
1977	60,000

Source: EUSIDIC survey 1977 and Ulrich's International Periodicals Directory

FIGURE 1

In order to try to control and make the most effective use of the available information many companies have devised, or are now trying to devise, new ways of providing information services. Clearly, an automated information retrieval service represents a particularly promising way of providing an information service for management.

B The Present Interest in On-line Information Retrieval Services

Several bodies in the UK are taking an active interest in on-line. For example, the British Computer Society has an Information Retrieval Specialist Group whose purpose it is to promote awareness of the different systems.

ASLIB (the Association of Special Libraries) which has a long standing interest in library developments, provides a full service of advice and search facilities. The Institute of Information Scientists provides an educational programme which places an increasing emphasis on international on-line information. The Library Association arranges courses, and is likely to become more active as public libraries start using on-line information facilities.

Of a more specialised interest are such organisations as the United Kingdom Chemical Information Service which provides a service in the field of chemistry. In the area of medicine, the British Library at Boston Spa has extensive knowledge. The universities of London, Loughborough and Manchester have been involved in on-line information retrieval systems for several years.

The on-line system operators provide regular newsletters which provide an essential service for the user from the point of view of new facilities, databases, charges, etc. Two new journals — *On-Line* (which is a US journal) and *On-Line Review* (which is a UK journal — give comprehensive coverage to latest developments, database comparisons, etc.).

The international body, EUSIDIC (the European Association of Information Services), has been formed by large organisations to influence and develop information services in Europe.

A recently formed EURONET User Forum meets bi-annually to discuss developments, and EURONET News is a newsletter which is issued from time to time.

The recent introduction by the British Post Office of access facilities to US databases is another example of the interest that has been generated in on-line information retrieval services. The demand for this service has already considerably exceeded all expectations.

C The Origins of On-line Information Retrieval Services

Automated information retrieval services first appeared in the United States when companies, such as Lockheed Information Systems and Systems Development Corporation, began marketing information (consisting mainly of journal abstracts and citations) that those companies had originally collected for their own use. Since then, a commercial market has developed for such information, and more and more companies have been, and still are being, set up specifically to sell information. Unlike the early pioneers (who gradually developed in-house information systems and then made them available to external customers) the new entrants to the market buy in their databases from database producers.

D The Main Elements of an On-line Information Retrieval Service

1. The difference between data processing systems and on-line information retrieval systems

There is an important difference between conventional data processing systems and purpose-built on-line information retrieval systems. Data processing systems — including those that use database management systems — can provide an adequate service if users' enquiries of the database are comparatively limited and infrequent. If, however, the nature of the enquiries demands more complex and more versatile search facilities, and if, also, a reasonably fast response time is required, then the information system needs to be designed

specifically to handle the storage and retrieval of information. However, future developments in hardware and software (including those discussed in the Section that begins on page 28) may reduce or even eliminate the difference that exists between these two different kinds of system at present.

2. **The filing system used in on-line information retrieval systems**
The primary objective of an on-line information retrieval system is that it should be possible to retrieve the desired information quickly and easily. To enable this to be achieved, most on-line systems use what is known as an 'inverted filing system'.

   A description of an inverted filing system is given starting on page 30. Briefly, however, the essential feature of an inverted filing system is that certain of the data items that are entered into the database are classed as keywords, and are entered as such into the index file. As each entry of this data item is made into the main file, its position is noted by allocating a reference to the entry in an index file.

3. **Bibliographic and factual databases**
The bulk of the information that is held in information retrieval systems is of a bibliographic nature. Bibliographic information is mainly textual, and it usually consists of either abstracts from, or citations of, publications. Bibliographic information is sometimes called 'secondary information'. The searching of a bibliographic database is just the first step towards obtaining the information that the searcher actually requires. The second step is that of obtaining the 'primary information' (the actual publication) through (say) a library or a bookseller.

   In the past few years, more and more databases which hold the primary information itself (factual databases as they are sometimes called) have been added to vendors' catalogues. Examples of the type of information that is held on these databases are:

   - Market research findings.
   - Economic information.
   - Company descriptions and statistics.
   - Product specifications.
   - Timetables.

   Often, these databases are numeric rather than textual in content. Because of this characteristic the vendor is sometimes able to enhance his revenue by providing users with facilities for manipulating the retrieved data.

4. **The searching of an on-line information retrieval system**
To search a database in the most efficient manner demands the services of an individual who possesses the necessary specialist skills. Because the present-day costs of on-line searching are high a highly-skilled searcher can represent a good investment. Consequently, some medium-sized and large companies are either recruiting, or are training, experts to undertake their searching work. To work with maximum effectiveness, the searcher must be able to appreciate what the user really wants out of the search. An additional attribute is a good knowledge and appreciation of the total company operation.

5. **The accessing of an on-line information retrieval system**
In order to access a database, the user requires a teletype-compatible terminal together with a communications device such as a modem or an acoustic coupler. Where a user has a requirement for browsing, a visual display unit and a printer should, desirably, be used.

6. Data networks

The data network is the medium through which the searcher communicates with the database itself.

Networks have evolved as the need for links between one computer and another, and between terminals and a computer, has grown. In many ways, the network is analogous to the operating system of a computer. The early computers had very little, if anything, in the way of operating systems, and, as a consequence, the programming of a computer was a complex task that often resulted in unreliable, and error-prone software. Later, as computers developed in complexity and scope, more and more of the common jobs of programming were built into the operating system. Today, computers have very high-level programming languages which are easy to use and enable very reliable programmes to be produced.

In a similar way, networks have taken the basic communication link of a telephone line, and have added value to it by providing error checking, line optimisation and many other features. The result of these 'value added services' (as they are called in the US) is that communication links are very much more reliable and much more cost effective.

There are important differences between networks in the US and networks in Europe. In Europe, each country's PTT controls the public data networks in that country. This does not apply in the US, where the government allows private organisations to develop networks and to offer a network service to the public. Two of the more well known of these private networks are TYMNET and TELENET.

The creation of diverse networks, with each one having its own design objectives and solutions, has produced compatibility problems arising from the need for computers and terminals from different manufacturers to communicate with one another. The Consultative Committee for International Telephone and Telegraphs (CCITT) have attempted to bring some order to the growing problem of laying down standards for suppliers. (A full discussion of compatibility is provided in Report Series No. 3 — Terminal Compatibility.) One important standard that the CCITT has laid down is the X25 interface for computers and terminal devices that are connected to packet-switching networks.

Packet-switching is the predominant technology for networks at the present time, and the general acceptance of X25 is bringing about a more coherent approach to network development. (A detailed discussion on networks is provided in Report Series No. 7 — Public Data Services.)

One aspect of packet-switching networks that is particularly important to on-line information retrieval is that of tariff structure. Data transmission and voice transmission through the public switched telephone network are both subject to a similar tariff structure, in which the cost is directly related to the distance over which the data is to be transmitted (i.e. the cost is distance dependent). This means that the user who is sited in the same location as the database that he wishes to access incurs lower costs than a user who is located at a remote site.

Data transmission through a packet-switching network is charged on the basis of how many packets of data are transmitted irrespective of the distance between the user and the database. This method of charging is called the distance independent tariff.

The majority of the public data networks that are being developed in Europe today are packet-switched networks. One of the most significant of these is EURONET. EURONET, which is a European Economic Community project, is discussed on pages 15 to 17. It has been developed specifically to allow organisations in member states of the EEC to have access to one another's databases, and it will probably grow to include both database

vendors and users from outside the EEC. EURONET provides the best example to date of the growth of on-line information retrieval systems in Europe.

The dominance of public data networks in Europe suggests that 'integrated system vendors', who provide a complete service encompassing database producer, database vendors, communication vendor and so on, are unlikely to appear in Europe.

E Viewdata

Viewdata, which is a comparatively recent, and a British, invention will undoubtedly affect the growth pattern of on-line information retrieval systems in both Europe and the United States. This report is not directly concerned with viewdata systems, but the Section beginning on page 24 briefly describes viewdata. It also explains how it can be expected both to compete with, and to complement conventional information retrieval services.

IV. ON-LINE INFORMATION RETRIEVAL SERVICES TODAY

A Introduction

Currently, database vendors in the US dominate the world scene. The three major textual bibliographic on-line retrieval vendors are Lockheed, System Development Corporation and Bibliographic Retrieval Services. The British Library's BLAISE system and the European Space Agency database in Frascati, Italy (both of which are discussed below), are two of the largest systems in Europe.

It is very difficult to estimate the size of the world market. Probably the most accurate available estimates for the US are those published by the Information Retrieval Research Laboratory in Illinois which estimates that in 1977 2,000,000 on-line searches were made on bibliographic databases.

In the US, the charge rate for information retrieval varies with the database. Prices range from $25 to $150 per connect hour, and it is reasonable to assume that the average is about $60 per connect hour. If an average search is assumed to take 20 minutes, the estimated value of the on-line bibliographic search market (excluding factual databases) for 1977 was around $40 million, exclusive of telecommunication charges.

In Europe, the most accurate available estimates are those published by EUSIDIC. EUSIDIC's most recent survey estimated that, by the end of 1976, there had been 200,000 on-line searches of bibliographic databases in Europe. The Technical Reports Centre service, Dialtech (the UK service to the European Space Agency — whose system is described on page 19) has estimated that the cost of an average search is about £10, excluding telecommunication charges. The value of the European market for on-line searches of bibliographic databases at the end of 1977 can, we consider, be reasonably estimated to be about £3,000,000, excluding searches of factual databases.

The EUSIDIC survey referred to above estimated that 2,000,000 factual databank searches were made during 1976. The Information Retrieval Research Laboratory estimates referred to above indicate that of an estimated world total of 362 bibliographic databases as at December 1977, 208 originated in the US. Of these 208 databases, about 75% contain technical information (covering subjects such as science, medicine and technology), and about 25% contain commercial information.

Many of the early US databases were created by organisations in the public sector, mainly in federal and state governments. It is interesting to note, however, that today only 23% of the on-line databases in the US are in the public sector, the remaining 77% being in the private sector. Also, 60% of the private sector databases have been specifically developed for profit.

The majority of European databases are in libraries, universities and government departments. In the UK, the major developments have occurred through funded research, such as that by the British Library, the United Kingdom Chemical Information Society, and the Institution of Electrical Engineers. The German Government are developing a well-planned series of databases with several centres, each of which will specialise in a different subject.

The European network EURONET (described on page 15) has received substantial public funding through the Commission of the EEC. A recent report which was produced for the Commission ranked the nine member countries of the EEC who are participants in EURONET in the following three groups:

- Germany, the UK and France are the most advanced in on-line information retrieval systems.
- Belgium, Denmark and Italy have an intermediate position.
- Ireland, Luxembourg and the Netherlands have made the least progress with on-line information retrieval systems.

Because of its connection to the Nordic network SCANNET, which covers Sweden, Finland and Norway, Denmark occupies a special position. Non-EEC members will almost certainly be interested in EURONET, and the Commission of the EEC have indicated that they will consider such participants.

Most of the participating countries are building their own national networks which will link into EURONET (e.g. the French TRANSPAC, the UK's PSS, and the Dutch DNI). (Report Series No. 7 gives a detailed report on Public Data Services.)

B Examples of On-Line Information Retrieval Services in the US

As with so many products, the US were the first country to establish a country-wide on-line information retrieval system, when System Development Corporation, in 1965, began operating an on-line bibliographic retrieval service ORBIT to service the whole of the US. In 1974 the service became an international one, and today the ORBIT system and the Lockheed DIALOG system are the world market leaders in on-line information retrieval systems. The Information Retrieval Research Laboratory's estimates clearly show the growth in the number of on-line searches in the US over the years 1974 to 1977, as under:

1974	–	700,000
1975	–	1,000,000
1976	–	1,200,000
1977	–	2,000,000

There have, of course, been both profitable and unprofitable databases in the US, and some of these are discussed below. The information there gives some insight into the different types of information that are currently available in the US.

1. **Market research companies**
   Some companies have realised that, as a result of conducting their main business activity, they have now produced a valuable database which they can sell as an on-line information retrieval system.

   Market research companies fall into this category. Traditionally, market research companies such as A.C. Nielsen and the Gallup Organisation, charge their clients a fixed annual fee, and supply them with information relating to their particular business needs. The availability of on-line databases will obviously have a dramatic impact on the business activities of those market research companies. It probably provides them with the opportunity to maintain their current business and also to add a new dimension in the form of the on-line retrieval of information. It is still too early to forecast what the outcome of this develop-

ment will be and, so far as is known, these companies have not yet developed firm policies regarding on-line information retrieval systems.

A. C. Nielsen have acquired a regional time-sharing company in Minneapolis. They have put up their databases on to the system, but initially they are allowing only a few carefully-selected clients, together with the Nielsen in-house research teams, to have on-line access to the database information. By studying the behaviour of this sample of users, A. C. Nielsen hope to be able to assess what the impact on their business will be when clients can select, on-line, specific items of information that they require for their particular needs.

The correct pricing strategy will, of course, be a critical aspect in A. C. Nielsen's future in the information retrieval industry. Equally important is how Nielsen's clients take to the service. If they find that the on-line system is easy to use, and that they can quickly obtain the concise information they want, then there will be a direct correlation between the use of the information and the use of the system. This correlation does not exist with the present methods, where the client buys, for a fixed fee, a large volume of information much of which he never uses. With an on-line retrieval system, the client will buy only the information that he requires. The remaining information will be available for any other client, and some of it may never be accessed by any client.

Interactive Market Systems Inc., who provide on-line and batch access to most of the information services of the large survey firms, realised early on the potential of this type of service. They entered the market in 1969 and, by taking over their competitors, have now achieved an almost monopolistic position in the supply of pharmaceutical information. As a result, they now occupy a very prominent position in the on-line field of market research. The Columbia Broadcasting System Inc. recently made an unsuccessful $100 million bid to acquire the company.

2. **Real estate**

Tymshare Inc. of California entered the field of on-line information retrieval systems in the early 1970s when they realised that there was a business opportunity in acting as the essential network link between database vendors and database users. They set up the first network, TYMNET, and they are now registered by the Federal Communication Commission as a common carrier. They now perform a very necessary function in the on-line industry.

Tymshare's interest widened in parallel with the growth of the network business. About four years ago they saw that they had an opportunity to develop an on-line numeric database on real estate information. With the development of their system, Tymshare have offered additional manipulation facilities, and the business has grown accordingly. Tymshare now have a well-established business in real estate information services in the US and, with TYMNET reaching out as it is now doing, Tymshare's entry into Europe and Japan can only be a matter of time.

3. **Company profiles**

During the last four years Dun and Bradstreet Inc. are said to have spent up to $20 million on computerising their credit files. They now have the facility to input new files into the system, and to retrieve data on-line. Like A. C. Nielsen, Dun and Bradstreet are now in a position to offer their service direct to the public. However, they are said to be nervous about doing so since they cannot judge whether, and to what extent, this will have an adverse effect on their existing business.

TRW Credit Services Inc. of California perhaps saw this hesitation on the part of Dun and Bradstreet and are now attempting to capture a share of the market. In conjunction with the National Non-profit Association of Credit Executives they have established an on-line facility that is available direct to the user. The system is not as comprehensive as Dun and

Bradstreet's, and perhaps as a result has not been as successful as TRW Credit Services probably hoped. However, the move indicates that the on-line business is becoming competitive. In the past, as mentioned on page 10, the majority of databases were set up and formed by non-profit organisations such as government departments. This venture by TRW Credit Services is interesting in that it merges the activities of both profit and non-profit organisations in order to produce an on-line service. This development may well presage other similar mergings of interests.

4. **Financial**
Telstat Systems Inc. was formed in 1969 by Penny Kaniclides who, seeing an opportunity, left McGraw Hill to form a new company that would provide an on-line access facility to financial databases. Bankers, insurance companies, brokers, etc. use the service, and Telstat's 1975 annual report showed an annual turnover of almost $3 million and a profit of about $1 million. These figures illustrate the high level of profitability that can be achieved in the on-line information retrieval business. Western Union Inc., probably because they were keen to break into the on-line information retrieval business, have now acquired Telstat Systems for a figure that is reported to be in the region of $15 million.

5. **Timesharing bureaux**
Tymshare Inc., mentioned on page 12, is not the only example of a timesharing bureau that has diversified by entering the on-line information retrieval business. Timesharing bureaux generally have had to look for other sources of income because many users of timesharing bureau systems have acquired one of the many small business computers that are now available. Some timesharing bureaux now supply a turnkey operation for small business computers, but others have entered the on-line information retrieval business.

One example of this latter trend is that ADP Network Services Inc. now offer the COMPUSTAT database (COMPUSTAT being the US equivalent of EXSTAT). This example is a most interesting one, because ADP Network Services are not the sole distributors of COMPUSTAT. Interactive Data Corporation (IDC) also provides the facilities of COMPUSTAT.

6. **Integrated systems**
One of the most successful on-line database operations in the US is that of Data Resources Inc. The company was founded in 1969, and at the end of 1977 its balance sheet showed a turnover of $24 million and a profit of $2.5 million. Several factors have contributed to the dramatic rise of this company. It is an integrated systems vendor, which means that it acts as a database producer, a database vendor, and a communications vendor. It operates its own network, but it also uses TYMNET and TELENET to provide a total service to over 500 clients. The company's mix of in-house and bought-in databases is claimed to be the largest collection of economic databanks in the world.

As well as providing on-line access to information the company also provides:

- Data manipulation facilities (e.g. modelling and forecasting).
- Training and education.
- Consultancy in business problems.
- Conferences.

The company's major source of revenue is from clients who use the computer for a variety of purposes. The company has, therefore, geared all its services in such a way as to ensure that they all involve the use of the company's computer.

The company has successfully introduced the concept of an annual subscription package.

It offers 16 packages, and each package gives comprehensive coverage to the application on which it is based. The annual subscription for the packages ranges from $5,000 to $25,000. The company's success is demonstrated by the fact that the annual report for 1977 showed that the annual revenue per client grew from $15,625 in 1971 to $42,692 in 1977. Any company that is contemplating a similar venture could do worse than learn as much as possible about the policy that Data Resources Inc. has adopted.

7. The New York Times Databank

The New York Times Databank is probably one of the best known examples in the world of an integrated system vendor. To date, however, the operation has not been a profitable one. The company lost a little over $1.5 million in 1976, and despite an improved marketing strategy and an increase in turnover in 1977, lost almost $1 million in that year.

Unlike some other large database operators, who went on-line as a development from hard copy publishing, the New York Times started from scratch, and its database was compiled at enormous cost without any accompanying hard copy revenue. It adopted a unique system which, by its nature, prevented the database being put up on other on-line systems, and so limited the degree of market penetration.

The early development costs have been estimated to have been at least $12 million. Meantime, the company has changed its original policy of selling the databank as a research tool for the use of libraries, and now sells it as a decision-making tool for the use of business organisations. Only recently has it started to sell reports produced through the databank.

Other publishing organisations appear to be moving cautiously into this field. They are, no doubt, aware of the comparative lack of success the New York Times has achieved so far, and, almost certainly, are anxious to avoid making the mistakes the New York Times has made.

8. ADP Network Services Inc.

As already mentioned on page 13, ADP Network Services Inc. have been successful in selling COMPUSTAT. They were not successful, however, when they put up a chemical information system which concentrated on mass spectrometry. Their failure can be attributed mainly to the fact that such a specialised subject would appeal only to a very limited market of research workers with limited budgets. A contributing factor, however, was that the salesmen who were selling the service did not understand the subect area. The combination of these two factors meant that only a low volume of sales was achieved. Instead of an expected revenue of $1,000 per month, per customer, sales were, apparently, only about $150 per month, per customer. This venture perhaps illustrates that although on-line information retrieval services represent a potentially lucrative market, success is unlikely to be achieved without a proper, carefully-planned marketing approach.

C Examples of On-Line Information Retrieval Services in Europe

As already mentioned, Europe is lagging behind the US in terms of providing bibliographic retrieval facilities, and it seems that about 75% of database access at the moment is from the US. At present, the majority of European databases can be found in libraries, learned societies and government bodies.

One area of on-line information retrieval where Europe holds a leading position is that of viewdata, which is described in the Section that begins on page 24.

It is to be hoped that Europe will take advantage of its late start and learn from the mistakes that organisations of all kinds have made in the US. However, so far as on-line information retrieval services are concerned, there are two major differences between the US and Europe. The first is the regulatory position that the PTTs in Europe adopt, and the second is the

language barriers that exist between the different countries in Europe.

The regulatory position of the PTTs in Europe may well inhibit the successful introduction of an entrepreneurial venture such as that of Data Resources Inc., which is referred to on page 13. On the other hand, however, it may well produce a more coherent set of public data networks that give extensive coverage to all parts of Europe. The joint EEC venture of EURONET described below is an encouraging step in this direction.

The Commission of the EEC is considering the language problem brought about by the connection of international links of the following three fronts:

1. Automatic and semi-automatic language translation.

2. Terminology databanks.

3. Multilingual thesauri.

As on-line information retrieval services continue to grow, both in scope and in number, the issues of privacy and security will increasingly demand attention.

The following paragraphs discuss both EURONET and the most prominent database vendors in Europe.

1. **EURONET**
   The European User Related and Oriented Network (EURONET) will provide on-line access to computer-based information services to the EEC. The EURONET effort is being directed by the Directorate General XIII of the Commission of the EEC. EURONET has been under development since 1971, and its implementation is scheduled for August 1979.

   The EEC's declared objectives of EURONET are "the development of a community-wide network", and "voluntary integration of services into a community-wide information utility". The three laid-down guiding principles during the development of EURONET have been:

   a Access to the database is to be equal to all users, regardless of their geographical location, nationality, etc.

   b Command languages and protocols are to be standardised.

   c Competition is to be reasonably free, but may be moderated by the need to avoid excessive duplication, with subsidised services permitted only to the extent that they do not threaten to inhibit competition.

   The network is based on the use of packet-switching technology. It will have four switching nodes, located respectively in Frankfurt, London, Paris and Rome, and the four nodes will be interconnected by 48-kilobit lines. The remaining member countries will have terminal access facilities located at Amsterdam, Brussels, Copenhagen, Dublin and Luxembourg. A standard CCITT X25 interface will be used for connecting host computers and packet terminals. Non-packet terminals will connect to the network via a PAD (Packet Assembler/Disassembler).

   The financial commitment has been spread over six years, with £5 million being allocated to the first action plan (1975 to 1977) and about £7 million being allocated to the second action plan (1978 to 1981).

   As a general rule, database vendors (the host operators) must pay their own connection cost. However, the Commission is funding the pioneer connection work for the following

three hosts:

1. The German Deutsche Institute fur Medizinische Dokumentation und Information (DIMDI) (which uses a Siemens 7.755).

2. The European Space Agency (which uses an IBM 370).

3. The University of Rome (which uses a UNIVAC 1100).

Two other systems will make up the initial network for the EURONET launch:

1. The British Library's BLAISE system (which uses an IBM 370).

2. The Commission's Luxembourg system (which uses a Siemens 7.740).

As at February 1978 the number of host operators was as follows:

— France	8
— West Germany	6
— Italy	5
— UK	2
— Denmark	1
— Luxembourg	1
— Belgium	1

When the 24 hosts are on-line, it will be possible for users to access 98 different databases from several different countries. The 98 databases are made up of 76 bibliographic reference databases and 22 factural databanks.

A survey of European databases which ASLIB and EUSIDIC carried out jointly in 1977 showed that there were 486 machine-readable databases across Europe. Of these, 337 were bibliographic databases and 149 were factual databanks.

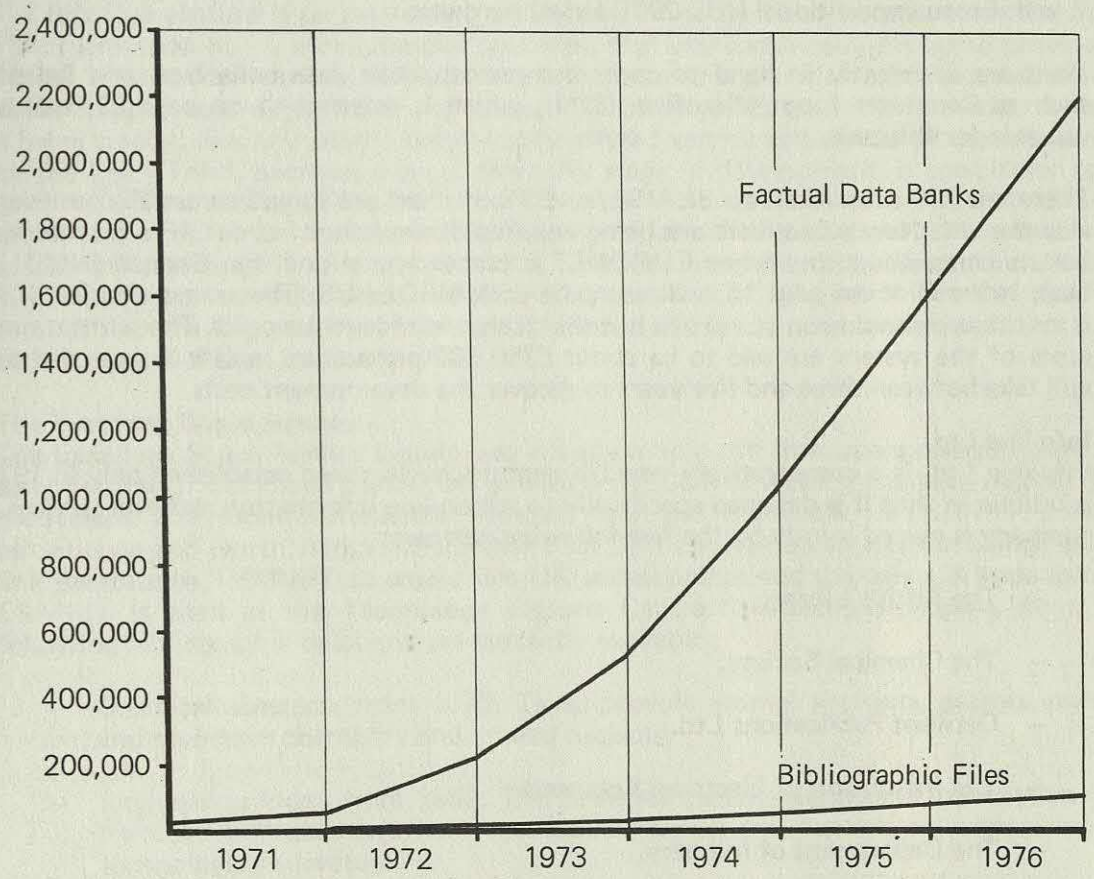
The graph (reproduced as Figure 2) from the report that was produced after the survey shows that, although they are fewer in number, factual databanks are accessed at a much higher rate. The Commission of the EEC is said to be aware of the greater potential profitability of factual databanks, and to be hoping that, as the network develops, additional databanks will be put on to EURONET.

This aspect of databases is only one of the many aspects that are being brought to light as EURONET develops. Other aspects that require discussion and research are:

- The development of a common command language.
- The multilingual aspects of databases.
- The technical aspects of document delivery.
- The copyright aspects of document delivery.
- Security and privacy.
- Tariffs.

When any large project becomes operational, it is rare to find that the early results match those that were expected when the decision to embark on the project was taken. It is appropriate to remember this as the day of the EURONET launch comes nearer. Many potential users of EURONET are now sceptical about its chances of success. As is well known, the original proposed launch date of January 1979 has been postponed. As against this, the announced EURONET tariffs are considerably less than those for certain of the US networks, such as TELENET and TYMNET. It remains to be seen, however, whether the European database vendors can match the service that both Lockheed and System Development Corporation give in the US.

The European Space Agency system has been available for some time in the UK through Dialtech in Orpington, but users apparently prefer the superior facilities of the US systems. Many users in the UK are optimistic that the Info-line system referred to on page 18 will prove to be an attractive alternative to the systems that Lockheed and System Development Corporation provide in the US.



Source: EUSIDIC Survey, 1st July 1977 — Searches of European on-line services

FIGURE 2

Several recent studies have been carried out on behalf of the Commission of the EEC, and these predict that over half the database usage in Europe will be concerned with the two areas of medicine (36%) and chemistry (18%). The ASLIB/EUSIDIC survey referred to earlier indicates that there is a strong bias towards commercial applications. This evidence, taken together with the fact that factual databanks have a higher usage rate than bibliographic databases, gives grounds for believing that the Commission of the EEC has produced the wrong balance, and that commercial databanks will represent the major growth area with EURONET.

2. **The British Library**

The British Library has been extensively involved in on-line information retrieval for many years. The BLAISE (British Automated Information Service) system, referred to on page 16, went live in April 1977, and it is now probably the fifth or the sixth largest bibliographic database in the world. It is designed to enable its users to find out what has been published in the UK in any field of knowledge. MARC (Machine Readable Catalogue) files contain details such as title, author, and International Standard Book Number (ISBN), of all books that have been published in the UK since 1950. MEDLARS (Medical Literature Analysis and Retrieval System) files contain details of publications in the bio-medical field dating back to 1966.

The current system runs on the IBM 370 that is operated by the Rank Hovis McDougall bureau, and it uses the ELHILL software of the National Library of Medicine. The British Library is developing a new software system, MERLIN (Machine Readable Library Information). This software is designed for use with parallel processing technology, but initially it will run on conventional (ICL 2900 series) hardware.

Plans are apparently in hand to carry out retrospective data collection, and technology such as Computer Input Microfilm (CIM), which is referred to on page 29, will be invaluable for this task.

There are 320 subscribers to BLAISE, and 75 of these are spread across 15 countries outside the UK. New subscribers are being enrolled at the rate of about 15 a month. MARC has no competitors, but when EURONET becomes operational the German DIMDI database, referred to on page 16, will compete with MEDLARS. The stated aim for BLAISE is not that it should run at a profit but that it should recover its costs. The current running costs of the system are said to be about £750,000 per annum, and it is expected that it will take between three and five years to recover the development costs.

3. **Info-line Ltd.**

Info-line Ltd. is a comparatively new UK venture, having been established only in 1976. It is unique in that it is designed specifically to sell on-line information retrieval services. The company is owned jointly by the five following partners:

- The British Library.
- The Chemical Society.
- Derwent Publications Ltd.
- The Institution of Electrical Engineers.
- The Department of Industry.

The company has a competitive pricing policy. It claims that its databases are more accurate than those of other database vendors, and to substantiate this claim it is engaged on the time-consuming exercise of checking, and amending where necessary, the contents of all its databases.

It has so far announced the following three databases:

- Information Services for the Physics and Engineering Communities (INSPEC) Science Abstracts.
- The Chemical Abstracts together with the Compounds Registry Index.
- The Derwent Publications Ltd. World Patents Index (WPI).

Of the above three databases, the first two are scheduled to be in service during the first quarter of 1979, and the third is scheduled to be in service before the end of 1978. Info-line plan to have twelve databases on-line by 1980, and these will include BIOSIS and COMPENDEX.

Info-line is due to announce shortly a unique database, the European Chemical Business Database from Derwent Publications Ltd. This will provide economic and business information to the chemical industry.

Info-line's system software is a modified version of BASIS, which is an information retrieval system designed and used by the Battelle Research Institute of Canada. The hardware used is the CRC bureau's Univac 1100. There are at least ten local access points in the UK, and several in Europe. The advent of EURONET will widen Info-line's potential user base still further.

The Info-line venture is an interesting one for three reasons. First, because it is not a 'spin-off' organisation but is a commercial operation that was specifically set up to provide an on-line information retrieval service. Second, because the provision of user satisfaction is stated to be a main objective of the system. (In other words on-line information retrieval is being treated like any other commercially-offered service and not as an esoteric service for the few.) Third, because, even at this early stage of development, it can be seen to be based on a clear strategy. The European Chemical Business Database (which is referred to above, and soon to be introduced) has the chemical industry as its target market, but it widens the user base by aiming at business functions other than the research and development function. It would seem reasonable to expect that Info-line will adopt a similar approach to the engineering industry, and that it will branch out into new subject areas once it has established a reliable system.

4. **The European Space Agency**

The European Space Agency system was initially a 'spin-off' from space research work. An early version of Lockheed's DIALOG software (RECON-DIALOG) has been used to drive the system. The system is situated at Frascati, Italy, and is available, via ESANET, throughout Europe and North Africa. The British Post Office prevented an attempt being made to link up through TYMNET to create five UK access points, and the only UK node through ESANET is sited at the Technology Reports Centre (Dialtech), Orpington, Kent. The following bibliographic databases are currently available:

- Chemical abstracts from 1972. These provide journal abstracts, patents, reviews, and reports in chemistry and related subjects.
- Engineering index from 1969. This provides the engineering and information communities with abstracts from over 3,500 journals and papers from engineering and technological literature.
- Metal abstracts from 1969. These provide comprehensive coverage of international literature on science and applications of metallurgy.

A recent addition to the system is a databank of electronic components. This factual databank should prove very useful to design engineers, buyers of electronic components, and all who require information on electronic components.

Many of the European Space Agency databases can be found on the US systems and, although it is easier and less costly to access the European Space Agency than it is to access US vendors, the system has yet to achieve the success of similar systems in the US. It would seem that potential users are unhappy about the inferior quality of the software facilities, such as the comparatively long time a search takes, and the limited search capabilities the system has.

Generally speaking, however, reports can be produced with less delay than with the US systems. Recently, considerable effort has been made to improve the service (e.g. by the provision of new hardware and improved software) in readiness for EURONET. At the same time, the charges for the service have been increased, and are now comparable with those of other systems.

5. **The National Computing Centre (NCC)**
The NCC have been running an information retrieval system for more than ten years. The system software THOR (which runs on an ICL 1905F) provides computer users with information such as:

   - Details and descriptions of available hardware.
   - Details of available training courses.
   - Abstracts from journals.
   - Details and descriptions of available software.

The present mode of operation is via post and telephone queries, but the NCC are said to be interested in increasing the availability of the databases by providing on-line access facilities.

The NCC have stated that they intend to be a host computer for EURONET and that they will make their databases available to other vendors. They appear to be interested not only in EURONET, but also in viewdata.

6. **Multilingual aspects of information retrieval systems**
One of the many issues that EURONET raises is the problem of multilingual databases. The Commission of the EEC have funded a study into the development of a multilingual thesaurus. It is relevant to mention here that the ASLIB/EUSIDIC survey of European databases showed that of the 337 available bibliographic databases in Europe:

   - 76% have English as the carrier language.
   - 10% have German as the carrier language.
   - 10% have French as the carrier language.
   - 4% are in languages other than English, German and French.

Of the 4% of databases that are in other languages, four are respectively in Russian, Japanese, Czech, and Swedish, and seven are multilingual databases. Only four of the latter seven can be regarded as being really multilingual. These four are:

1. Titus
   Titus is a commercial database of information related to the textile industry. It is designed for use by textile managers, engineers and technicians, and it is prepared by the Institut Textile de France. It is available in French, English, German and Spanish.

2. IRRD (International Road Research Documentation)
   IRRD is a database covering worldwide literature and reports on all aspects of road transport. The database is prepared by the Organisation for Economic Co-operation and Development (OECD) of Paris, and it is available in English, French and German.

3. CANCERNET
   Cancernet is a database covering cancer and related sciences (immunology, virology,

etc.). The database is prepared by the Institut Gustave Roussy in France, and it is available in French, English, German, Czech, Polish and Spanish.

4. CREDOC
   CREDOC is a legal database prepared by the Centrum Voor Rechtsdocumentatie in Belgium, and it is available in Dutch and French.

Studies are being conducted by the EEC on multilingual databases, and it is likely that further multilingual databases will be developed in the near future.

D Examples of On-Line Information Retrieval Services in Japan

In any study of on-line information retrieval systems recent developments in Japan are relevant. In 1976 the Japanese Economic Journal reported on a market survey of on-line information retrieval systems in Japan. This predicted that the Japanese investment in information systems would grow from US $1500 million in 1974 to US $9000 million by 1985, representing an annual growth rate of 17.4%.

1. **Pattern Information Processing Systems (PIPS) Project**
   PIPS is a national research project which is sponsored by the Ministry of International Trade and Industry (MITI). Continuous research has been undertaken since 1971 and the budget for the project is US $175 million. It is hoped that the project work will result, in the early 1980s, in the development of a practical pattern information processing system.

   In the work done so far, pattern recognition research has been carried out in the following areas:

   - Printed Chinese characters.
   - Monochrome pictures.
   - Colour pictures.
   - Three-dimensional objects.
   - Voice.

   Much of the technology which is referred to in the Section beginning on page 28 is being developed for this project. For example, picture recognition is being implemented through parallel processing techniques. All the major manufacturers in the respective subject areas have been involved in the project.

2. **The Hi-OVIS (Higashi-Ikoma Optical Visual Information System)**
   In July of this year, one of the world's first visual information systems employing optical communications techniques went live on a pilot basis. The system, on which so far $15 million (US) has been expended, uses terminals that have been installed in 168 homes.

   The system has two-way communications facilities with a camera giving the centre access to each home. In each home, a microphone and a keyboard linked to the TV set gives the home access to the centre. The householder can retrieve information by means of his keyboard and TV set.

   Unlike a similar experiment in Tama, Japan, which was financed by the Posts and Telecommunications Ministry, the Hi-OVIS experiment is backed by the Ministry of International Trade and Industry (MITI). In backing the experiment the Ministry has taken

a long-term view of the manufacturing potential for new technology.

Three of the most advanced companies in their respective fields have been involved in the experiment:

- Fujitsu Ltd. (for the computer technology).
- Sumitomo Ltd. (for optical fibre transmission technology).
- Matsushita Electric Industries Co. Ltd. (for the audio-visual technology).

If the experiment is successful, the present plans are to extend the system to embrace a new town which is being built at Kobe, near to Higashi-Ikoma. If this happens, Kobe will be the first town in the world to be built up around a communications system. Also, if the experiment proves that it would be feasible, community systems could be set up throughout Japan, with information common to several communities being held in one central databank. Only three or four central databanks would be required for servicing community systems throughout Japan.

3. JICST On-line Information Service (JOIS)

JICST (The Japan Information Centre of Science and Technology) was established in 1957 by the Japanese government. Its purpose and activities cover the collection and the dissemination of scientific and technical information.

Since 1978, the JICST on-line information service has offered on-line access to Chemical Abstracts and MEDLARS (both of which are referred to on page 18), together with JICST's own abstracts database. The system is accessible throughout Japan via the JICST on-line network.

4. Nikkei Economic Electronic Databank Services (NEEDS)

NEEDS is one of the many information services offered by Nikon Shimbun Incorporated. The company was established in 1876, when its only product was the Nikon Keizai Shimbun (the Japanese equivalent of the Financial Times). The company has always maintained a keen interest in the information industry, and currently is involved in:

- Publishing newspapers, magazines and books.
- Publishing newsletters.
- Providing conferences.
- TV and radio broadcasting.
- Providing a market quotation service.

The company's gross revenue for 1977 was about US $300 million, which represented an 8.4% increase on 1976.

The 'Nikkei', as the Nikon Keizai Shimbun is known, is regarded as a total economic information system. All editorial material used for the newspaper is stored in the computerised databank. Since March 1978, by using ANNECS (Automated Nikkei Newspaper Editing and Composing System), the printing of the newspaper has been completely automated, and copies are printed simultaneously, through a facsimile network, in five cities. The newspaper has a daily circulation figure of 1.77 million, which is one of the largest in the world. (The Wall Street Journal has a circulation of about 1.54 million, and the Financial Times has a circulation of about 180,000.)

The company has formed links with some of the major organisations that are engaged in providing economic information, as follows:

- McGraw-Hill Inc.
- Dow Jones and Co. Inc.
- Reuters Ltd.
- Data Resources Inc.
- SVP of France.
- FIND/SVP in the US.

The company's move into on-line information retrieval was a natural step in providing a total information system. The databank, which is being held on a Burroughs 6800, provides both bibliographic and numeric information. The bibliographic database consists of reports from all the publications, and the numeric data consists of up-to-date economic information. Through its links with both Data Resources Inc. and Standard and Poor (McGraw-Hill), the system has acquired major users in North America.

Currently, about 80 clients use the on-line facilities, about 90 use the batch facilities, and about 100 buy the hard copy service.

5. **Developments in facsimile transceivers**
Facsimile transceivers do not form part of an on-line information retrieval service, but the Japanese experience in this field is sufficiently interesting to warrant the following brief account being included here.

Because of the pictographic nature of their language and its incompatibility with telex, the Japanese have always been well advanced in facsimile hardware. Until now, however, the regulatory control of the television airwaves has meant that the development has been limited to transmission over telephone lines. However, the manufacturers concerned are now optimistic that the controls will be relaxed soon, and they also hope that permission will be given soon to use the television airwaves. Facsimile transceivers are in widespread use for intracorporate communication, and the communications satellite that is proposed for information transmission throughout Japan could provide every Japanese home with the facility (through 'homefax') to access local and national databanks.

Matsushita Electric Incorporated (National Panasonic) have developed a combined TV set and ink-jet facsimile receiver. If this system were to use a high resolution colour tube, it could produce printed output of the highest quality.

V. VIEWDATA

A Introduction

When considering on-line information retrieval systems in Europe viewdata cannot be ignored.

Viewdata is the generic term for a public information service which uses a modified television set as a terminal device, and the standard telephone line as the communicating link to either the computer that holds the database or another terminal device.

Viewdata was invented by the British Post Office and initially it was the name given to the British system itself. However, viewdata became a generic term and so the name Prestel was given to the British viewdata system.

As already mentioned, viewdata is capable of providing more than on-line information retrieval. It can communicate with other terminals and through it, people with terminals can communicate with one another. Through this facility, people can "shop by television".

B The Status of Viewdata Today

Having invented the viewdata system, the British Post Office is currently in a strong position to take advantage of a rare marketing opportunity. The Post Office has already sold the system in one form or another to the PTTs of West Germany, Holland and Hong Kong, and is actively engaged in selling viewdata throughout the world.

Prestel is currently undergoing a market trial, and the public service is scheduled to become operational in the first quarter of 1979.

Since more than 150 information providers have contributed to its launch, there is clearly no shortage of interest in Prestel. A recent market survey carried out by a major manufacturing participant in Prestel, predicts there will be a cumulative growth in Prestel terminals in the UK, as illustrated in Figure 3.

Year	*Cumulative No. of Terminals*
1979	33,000
1980	108,000
1981	254,000
1982	587,000
1983	1,141,000

Figure 3 Cumulative Growth Estimates for Viewdata Terminals

It was originally thought that members of the public would represent the major market for Prestel. Now, however, it is realised that business and commerce represent a major potential market.

The Independent Television Authority and the British Broadcasting Corporation respectively have introduced teletext systems known as Oracle and Ceefax, and these are both often mistakenly considered to be viewdata systems. There are, however, significant differences between these systems and Prestel. They provide a limited information retrieval service, not via the telephone line, but via the television airwaves. Each of the systems simply transmits continuously on a cyclic basis, a fixed amount of information, which is restricted in scope in order to allow an acceptable response time to be achieved. The user is able to use a keypad to request a particular page of information to be displayed on the TV screen.

By contrast, there is no theoretical limit to the number of pages of information that a user can select from a viewdata database.

Outside the UK, France has probably the most technically advanced competitor to the viewdata and teletext systems with a system called Antiope/Titan. Antiope is the equivalent of teletext and Titan is the equivalent of Prestel. Antiope/Titan can provide a more flexible display than that of Prestel. However, this facility requires extra memory within the television set to handle the differences, and so the Antiope/Titan terminal devices will probably be more expensive than the Prestel terminal devices.

Recently, there have been meetings between the CCITT and the European Conference of Posts and Telecommunications (CEPT) to try to agree a European display standard, but standardisation is probably still at least three years away. Outside Europe, similar systems to viewdata are being developed. For example, the Japanese are working on a system called CAPTAINS (Character and Pattern Telephone Access Information Network System), and in Canada the Videotex service is being developed to allow data to be transmitted through normal telephone lines, modified cable TV circuits, and optical fibre links.

C Viewdata in the Context of On-line Information Retrieval Services

When considering the likely development of the European on-line information retrieval market, it is not possible to ignore viewdata systems.

The more deeply one considers viewdata the more one can see additional applications of it. Clearly, there is tremendous potential for the use of a single terminal device which is capable of retrieving information from both viewdata systems and also large databases, such as those provided by Lockheed Information Systems. For example, a simple viewdata system could act as a primary index, which the user would search prior to searching the larger, more comprehensive database.

There is also considerable potential for using viewdata to provide low-cost in-house information retrieval systems. As evidence of this potential, Whitbread & Co. Ltd. has started to develop an on-line in-house management information retrieval system using viewdata terminals.

A bigger and more sophisticated information system is being planned based on the use of viewdata terminals in the Stock Exchange's Teletext Output of Price Information by Computer (TOPIC). This system will provide users with up-to-the-minute share price information and news and it will eventually link into other on-line information systems, such as Reuter's Financial Information. It provides a clear illustration of the synergy that exists between viewdata and traditional information retrieval systems, which could play a vital role in the development of on-line information retrieval systems.

One of the major selling points of viewdata is that it is simple to use. (The operation of a viewdata terminal is comparable with the operation of a telephone or television set and can be learnt in about the same amount of time.)

COMPARISON OF VIEWDATA AND TRADITIONAL INFORMATION RETRIEVAL SYSTEMS

	VIEWDATA (PRESTEL)	TRADITIONAL INFORMATION RETRIEVAL SYSTEMS	
		BIBLIOGRAPHIC	FACTUAL
EXAMPLES OF INFORMATION PROVIDED	1. News 2. Advertisements (jobs and products) 3. Games and puzzles 4. Buyers' guide	1. Scientific 2. Technical 3. Pharmaceutical 4. Agricultural	1. Scientific 2. Technical 3. Pharmaceutical 4. Legal
CONVENIENCE FOR USER	Easy to operate	Requires a skilled searcher	Requires a skilled searcher
SEARCH FACILITIES	Select page number from a menu	Search using text strings, to obtain reference. Follow up reference through library	Search using text strings to obtain required information
MANIPULATION FACILITIES	No manipulation facilities	No manipulation facilities	Many systems offer the user a language facility to manipulate the information
NECESSARY EQUIPMENT	Modified TV set with an integral modem	Teletype-compatible terminal plus modem or acoustic coupler	Teletype-compatible terminal plus modem or acoustic coupler
HARD COPY FACILITIES	Limited facilities	Quality dependent on printer used	Quality dependent on printer used
GRAPHICS CAPABILITIES	Primitive	Not applicable	Quality dependent on VDU used
COLOUR	Seven colours or black and white dependent on the TV set	Normally black and white	Normally black and white
TYPE OF USER	1. Business 2. Home	1. Business 2. Professional 3. Research	1. Business 2. Professional 3. Research
OTHER FACILITIES	1. Message facilities 2. Transaction facilities	Ordering of source material	Ordering of source material
COSTS	1. Local telephone call 2. Frame access charge	1. Local telephone call 2. Network charge 3. Volume charge 4. Database charge 5. Print charge	1. Local telephone call 2. Network charge 3. Volume charge 4. Database charge 5. Machine usage charge for manipulation 6. Print charge

FIGURE 4

As can be seen from Figure 4, the two types of system complement one another. Viewdata, although it is very much more limited in its capabilities, is a considerably cheaper method of information retrieval (see Figure 5).

SUMMARY OF COSTS FOR THE USER OF INFORMATION

Components of Cost	Viewdata (Prestel)	N. American Databases through TYMNET AND TELENET	European Databases through EURONET (Proposal)
Telephone charges	Local call rate	Local call rate	Local call rate
Database access	2p per frame (average)	£30 per hour (average)	Not available
Network charges	Not applicable	£3.30 per hour	£1.35 per hour
Volume charge	Not applicable	30p per thousand bytes	£1.15p per thousand segments (1 segment = 64 bytes)
Subscription charge	Not applicable	£5.00 per quarter per user name	Not available

FIGURE 5

The scenarios for viewdata are many and varied. However, one thing does look likely – viewdata will prove to be a very important factor in the growth of on-line information retrieval systems.

VI. THE FUTURE FOR ON-LINE INFORMATION SERVICES IN EUROPE

A Introduction

The on-line information services industry will grow steadily over the next five years. A recent report, based on a study of the US market by Frost and Sullivan Inc., forecasts that database-generated revenue will climb from $740 million in 1976 to $1,100 million by 1980, and then to $1,600 million by 1985. Clearly, these figures do not cover only the sale of on-line information services: they also include sales of database systems software products. However, according to Frost and Sullivan, "the sale of information contained in electronic database faces a future with almost no limits", and the fastest growing areas are:

Market Area	Annual growth
Business	25%
Econometric	20%
Legal research	20%

The report also states that more and more medium-sized and small companies are using the services of database vendors. It predicts a very bright future for on-line information retrieval services, with improved technology in data capture and data storage providing a cheaper and better service. The report forecasts that, because of cheaper memory costs, graphical databases (such as engineering drawings, maps, etc.) will become commonplace.

The recent survey of information systems in Japan by the Japanese Economic Journal, referred to on page 21, predicted that there will be an annual growth rate of 17%, and that the business in Japan will stand at around US $9,000 million by 1985.

The indicators in Europe, which are discussed below, suggest that there will probably be a similar rate of growth in Europe. As discussed on page 15, EURONET is scheduled to go live in August 1979, and this network will link all the EEC countries.

Info-line Ltd., discussed on page 18, is due soon to commence its operations in the UK.

B The Hardware and the Software Developments that are Likely to Affect The Future of the On-line Industry

As discussed earlier, on-line information retrieval systems are being developed across the world. Successful and profitable on-line information systems are already operating using today's existing technology. However, both the new technology that has been recently developed, and the technology that is currently under development, will almost certainly produce substantial improvements by way of reduced costs and better service to users.

Below, we examine the developments in hardware and software under the five following functions:

1. The capture of information.
2. The storage of information.
3. The retrieval of information.
4. The output of information.
5. The transmission of information.

1. The capture of information

Until now, information has mainly been abstracted by individuals (e.g. scientists) scanning documents visually, and then extracting the relevant information to be input to a computer system. An example of an organisation that uses this technique is the Commonwealth Agricultural Bureau in the UK, which supplies its database to all the large database vendors in the US and Europe. It has 14 abstracting units throughout the UK, and each unit prepares data input forms on which the abstracted data is entered. The forms are typed up by an external organisation using a font that is recognisable by OCR equipment. The database is updated twice weekly, and 150,000 abstracts are added annually. The system works well, and the majority of errors are said to be generated not by the OCR equipment, but by the typists in the external data preparation organisation. Increasingly, source data is being captured in machine-readable form as it is created (e.g. by the use of word processors, and by automation in publishing). This method makes it easier to produce a database, even though it creates some problems of compatibility and format. This abundance of information does, however, present some management and control problems.

This data capture technique cannot, of course, be applied to the enormous amount of information that was compiled before computers were developed, and that now needs to be included in some databases. The task of capturing this earlier information would be simplified if a suitable light pen existed which was capable of recognising different fonts as it scanned printed information. Such a pen would not solve the problem completely, however, since some earlier information would need to be transcribed before it was scanned.

A recently-announced system in the US, Stewart Warner's Datafax, is a hybrid OCR/fax device that is capable of scanning both text and graphics. The scanner employs microelectronic solid-state technology, and so is very much more reliable than electromechanical technology. The scanner uses standard OCR methods when reading characters, and switches automatically to fax techniques when it encounters graphics, signatures, etc. It is claimed that a 300-word document, when it is read by OCR only, will take only six seconds to scan. For hybrid OCR/fax matter, the scanner will take fifteen seconds to read the same document. The preliminary price for the scanner is about $12,500.

Computer input microfilm (CIM) probably represents the most significant technological development in the area of large-scale data capture. The only known available CIM system is the GRAFIX I product from Information International Grafix Ltd. In simple terms the methodology employed by the system is as follows:

1. The source documents (in any font or handwriting) are photographed.
2. The format or the indexing terms are specified, using an interactive data tablet and a VDU.
3. The film is scanned and then converted into machine-readable format.

The equipment (which consists of a camera, a film scanner, an image processor, and a

system control computer) is expensive, since it costs about $2 million. In the UK so far, the only user of GRAFIX I is the Department of Health and Social Security. The DHSS is undertaking a conversion programme from card files, and has found that the system is capable of processing 5.5 million characters per day. The British Library is believed to be considering using CIM for large-scale retrospective data capture in order to expand its database.

The University of Montreal, when it was preparing its DATUM information retrieval system, successfully used CIM to capture reports (printed in both French and English) of the proceedings of the Exchequer Court of Canada over the previous twenty five years.

2. The storage of information

a. Introduction

Before discussing information storage, it is relevant to consider what is meant by information storage. In brief, it is simply the physical means of holding files (traditionally on magnetic disc and magnetic tape) in a machine-readable format. With databases, however, information storage involves much more. In order to provide the required quick access times and the sophisticated search capabilities, it is essential to have well-structured files and good indexing techniques. Through the use of inexpensive microprocessors it is now possible to build cost-effective hardware that incorporates many of the features which, on today's existing systems, must be implemented through software.

b. Present-day hardware and software

Storage is generally regarded as being composed of the two separate parts of hardware and software. Hardware is selected on the basis of its access time capacity, its cost, and so on. When the hardware has been selected, the file structures and the indexing are designed and constructed to match the hardware.

Present-day systems use magnetic discs as the memory media for the on-line storage of information, and they use magnetic tapes for archiving and distribution. The table in Figure 6 shows that, in terms of access rate, magnetic discs cannot compete with the latest developments in solid-state memory devices (bubble and CCD). Bubble devices being non-volatile, seem the more likely to supersede magnetic discs.

Optical tape, with a packing density about 4000 times greater than that of magnetic tape, seems most likely to be used as the archiving and distribution medium of the future.

The majority of today's large textual databases use an inverted filing system which consists respectively of an index file and a document file. Figure 7 gives an example of an inverted filing system. The index file contains keywords, which have been selected from the document file. Each keyword has a set of pointers that reference each occurrence of the keyword in the original document file, and so enable searching by keyword to be quick and effective. Techniques such as word fragmenting (see Figure 8) and text compression (see Figure 9) can give increased search capabilities. They can also give economies in disc storage that can be as high as 30%.

Inverted filing systems have, however, the following two major drawbacks:

1. Effective searching is dependent on a good index.

2. A good index can be almost as large as the files themselves.

The problems of indexing have been a source of debate for many years. Indexing does not merely involve creating a direct 'mapping' between the keyword and the document. Different users who come from different backgrounds, and who are accessing

TECHNOLOGY		RANGE OF COST/BIT (cents)		RANGE OF ACCESS TIMES (ms)	
		MIN	MAX	MIN	MAX
RANDOM ACCESS MEMORY	BIPOLAR	2	70	5×10^{-6}	150×10^{-6}
	M.O.S.	0.35	3.0	70×10^{-6}	800×10^{-6}
CHARGED COUPLE DEVICE		0.01	0.35	0.03	0.15
MAGNETIC BUBBLES		⟨0.01	0.25	0.2	10
FIXED HEAD DISC		0.02	0.15	1	100
MOVING HEAD		0.002	0.02	10	500

The above table compares five different storage technologies in terms of cost and access times.

Source: 'Advances in Computer Memories' – K. Baker of Plessey Microsystems Ltd. in *International Systems* April 1978

FIGURE 6

the same database, will often employ a different vocabulary. In order to resolve this problem of vocabulary a thesaurus needs to be used and, with some databases, the thesaurus needs to have multilingual capabilities.

These indexing overheads, and the increased software complexity they bring with them, can often require as much disc capacity as the database itself. The point will soon be reached where information retrieval systems will be constrained by the limitations of the computer's disc handling capabilities.

Most of the existing bibliographic databases hold only citations or abstracts of publications, but some systems are capable of storing the complete texts of publications. However, a system that would enable complete publications to be stored and retrieved economically would require the use of new computer technology.

EXAMPLE OF PART OF AN INVERTED FILING SYSTEM

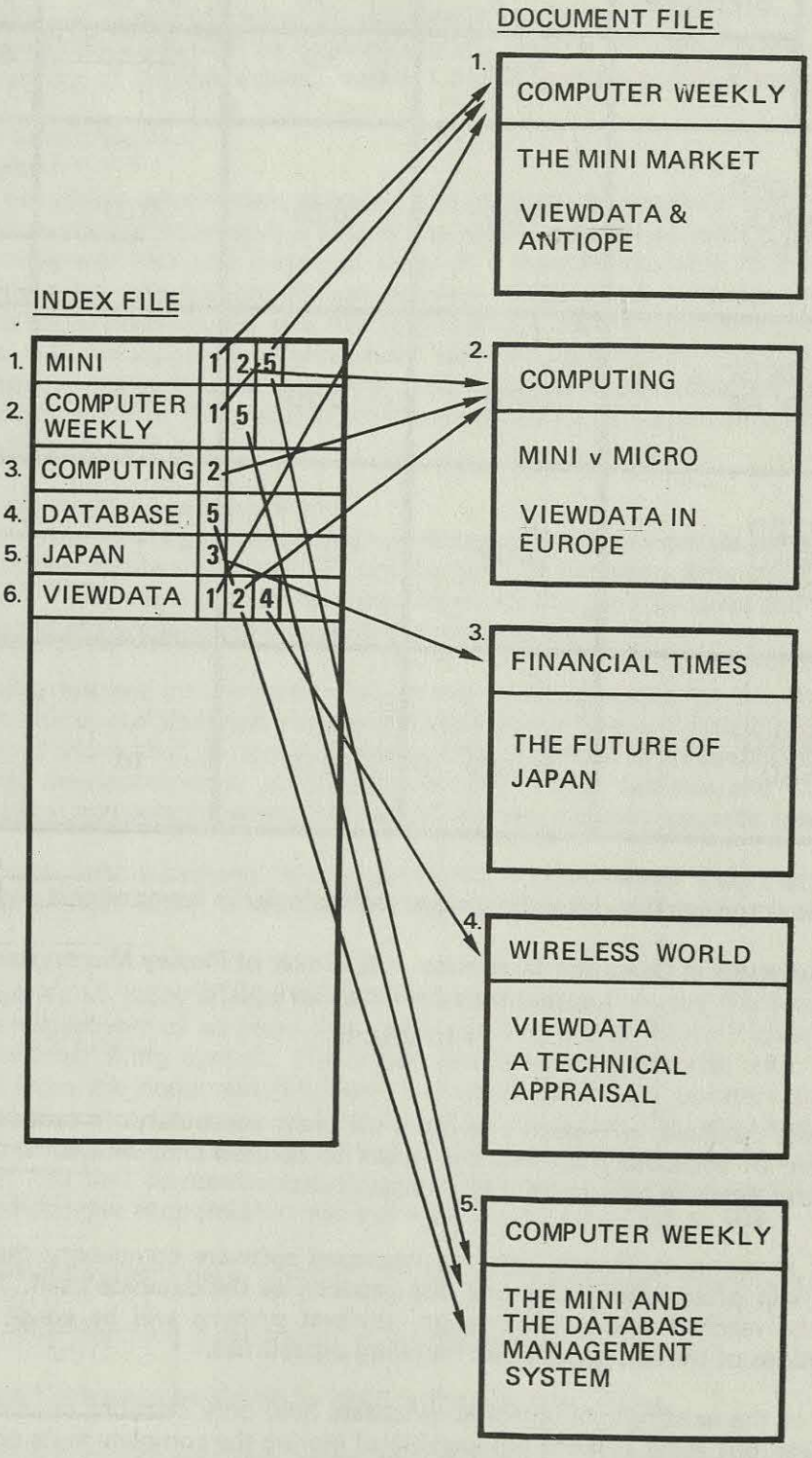


FIGURE 7

EXAMPLE OF PART OF AN INVERTED FILING SYSTEM USING A WORD FRAGMENTS INDEX

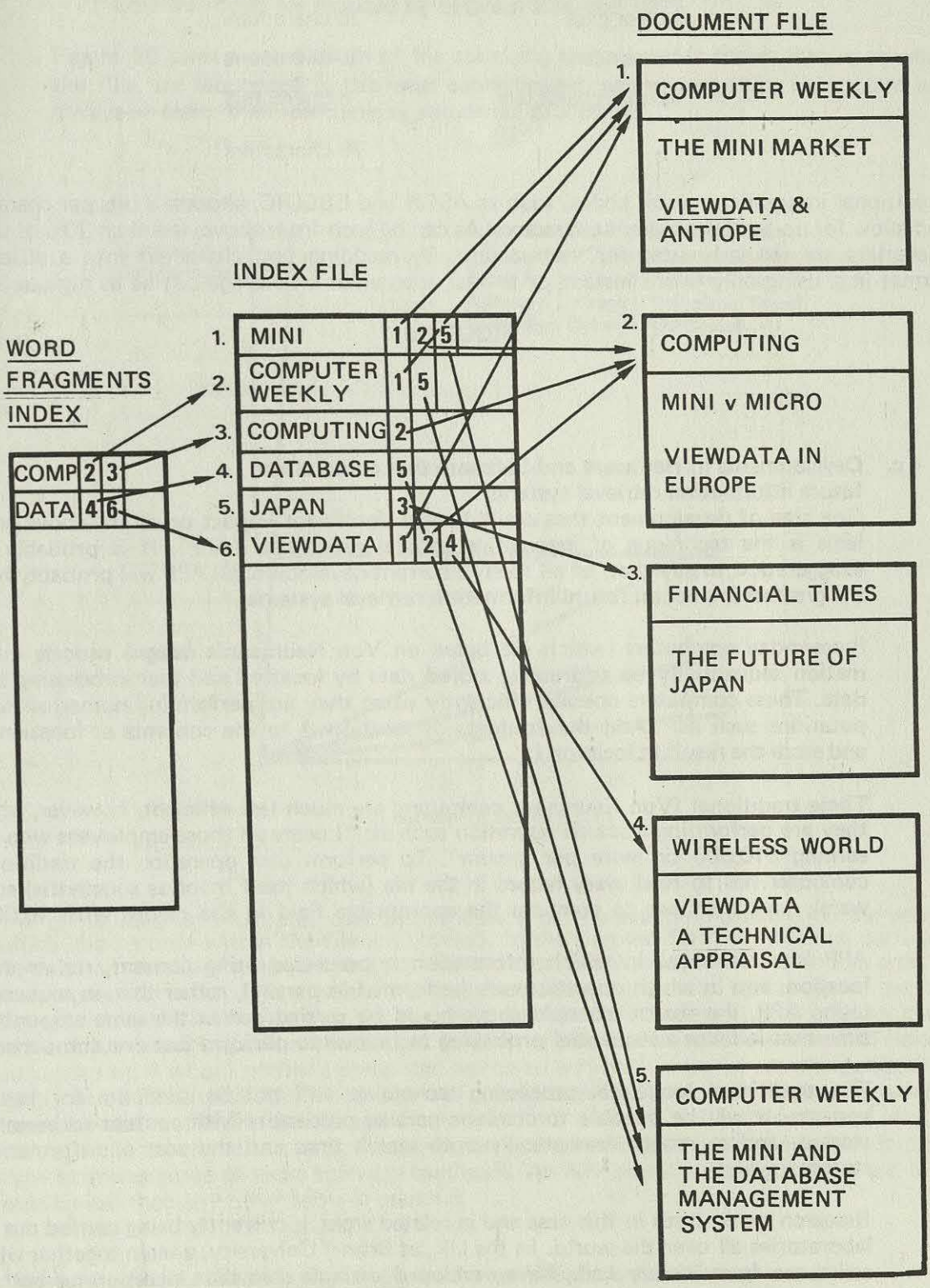


FIGURE 8

TEXT COMPRESSION

Upper case	26 characters
Lower case	26 characters
Numerics	10 characters
Special characters	8 characters
	70 characters

Traditional internal computer codes, such as ASCII and EBCDIC, allocate 8 bits per character and allow for up to 256 different characters. As can be seen from above, less than 30% of these characters are regularly used for text storage. By recoding text characters into a different format (e.g. using only 5 bits instead of 8 bits) economies on storage can be as high as 30%.

FIGURE 9

c. Developments in Hardware and Software that could affect future information retrieval systems

One area of development that could have a significant impact on all the above problems is the technique of associative parallel processing (APP). It is probably no exaggeration to say that, of all today's current developments, APP will probably have the greatest impact on future information retrieval systems.

Present-day computers (which are based on Von Neumann's design) process information sequentially by addressing stored data by location and then processing that data. These computers operate efficiently when they are performing numerical computations such as: "Add the contents of location A to the contents of location B, and store the result in location C".

These traditional (Von Neumann) computers are much less efficient, however, when they are performing a search operation such as: "Locate all those employees who are earning £10,000 or more per annum". To perform that operation the traditional computer has to read every record in the file (which itself involves substantial software), and then has to compare the appropriate field in the record with 10,000.

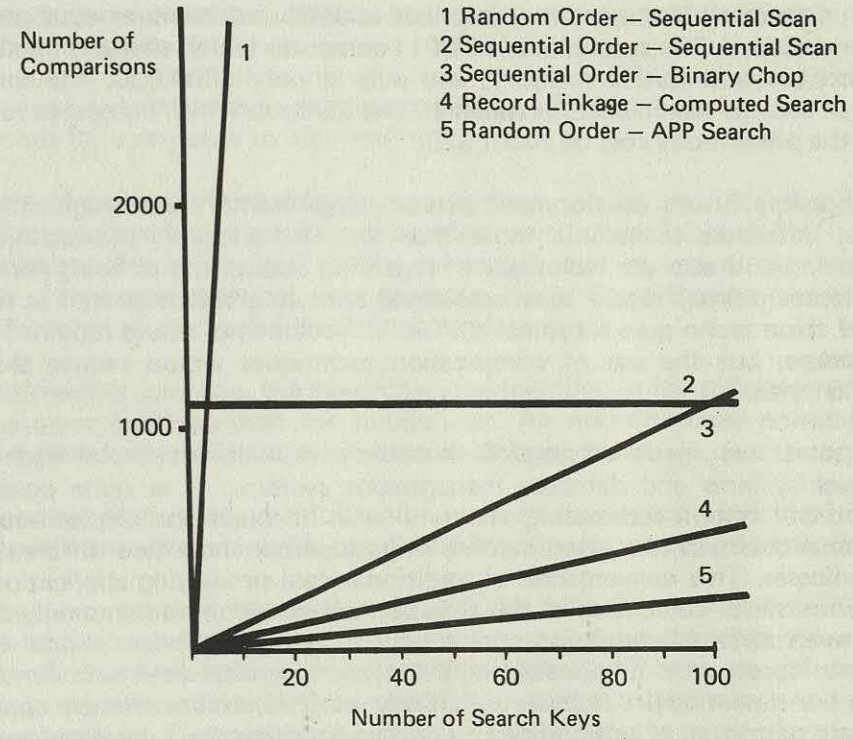
APP is a technique in which information is processed using content, rather than location, and in which operations are performed in parallel, rather than in sequence. Using APP, the search example above could be carried out in the same amount of time that it takes a sequential processing technique to perform just one comparison.

The traditional sequential processing techniques will not be adequate for future systems. It will be possible to combine parallel processors with content-addressable storage, and so reduce dramatically both search time and the cost of information storage.

Research work, both in this area and in related areas, is currently being carried out in laboratories all over the world. In the UK, at Brunel University, a team together with assistance from Plessey Ltd., have developed a single chip that incorporates both a content-addressable memory and a processor. Comparative research into information retrieval systems that use sequential processing by location has shown that APP produces the following improvements:

- Performance is improved by a factor of between 10 and 1000.
- Storage requirements are reduced by between 10% and 50%.
- Production costs are reduced by between 20% and 100%.

Figure 10 gives a comparison of file searching techniques. It shows that, even when the files are sequenced in the most sophisticated manner, searching by content will always be faster than searching by sequential processing.



The above diagram illustrates that the efficiency of file searching depends on the way in which the records within the file are ordered. In the simplest form of ordering, particular fields of a record are specified as sort keys, and records are sequenced in ascending order of their numerical value. This means that, if all the search keys are also sort keys, the 'sequential scan' curve 2 (assuming that the search keys are similarly ordered), or the 'binary chop' curve 3 file searching techniques would be fairly efficient. The addition of record linkage and indexing 4 would enable a computed search to achieve higher file searching efficiency but it would require additional storage and software. Where the format is based on a multi-keyword record the file is effectively unordered for most of the keywords, and very inefficient file searching results 1. File inversion and other indexing techniques can improve time at the expense of extra software overheads. An APP search is clearly quicker and more economical than any other form of search 5.

Source: Brunel University technical memorandum
"Associative Processing of Non-Numerical Information" – R.M. Lea

FIGURE 10 Comparison of File Searching Techniques

ICL have developed an intelligent disc subsystem (which is to be used initially with their 2950 DME and 1900 mainframes) which is known as content addressable file-store (CAFS MK2). ICL are aiming to sell the product to the owners of fast-response information retrieval systems. They claim that the system has a throughput 50 times faster than that of conventional systems, and, further, that it can support up to 500 users simultaneously.

The new software (MERLIN) that is being designed for the British Library BLAISE system (as mentioned on page 18 is based on APP, although it is expected to run initially on conventional hardware.

At the US Association of Computing Machinery (ACM) workshop, held in May 1977, on computer architecture for non-numeric processing, representatives of a US company, Operating Systems Inc., described a 1000 million byte database system they have developed. The system uses a PDP11 computer which employs parallel processors and content-addressable memory, and sells at only $600,000. The company claims that an average search costs between 10 and 25 cents which compares very favourably with the present-day cost of about $20.

The possible future development of very large memory with high speed access will enable databases to include more than text. By adopting the technique of image processing and also the technique of regarding documents as being formed of a grid of discrete points, future databases could contain graphics as well as text. With the use of these techniques a typical 8½" x 11" document would require 100,000 bytes of storage, but the use of compression techniques would reduce this to around 30,000 bytes.

The point was made on page 6 that there is a difference between information retrieval systems and database management systems. It is quite possible that the advent of content-addressable memory will, through the implementation of the relational database (discussed below), bring together these two different, yet similar, technologies. This convergence of traditional data processing applications and information retrieval could lead to the situation where terminals commonly communicate both with external databases and in-house databases. The relational database is a database constructed by considering the relationships between data items. The originator of the relational database, E.F. Codd of IBM, claims that by applying mathematical techniques of sets and groups to data handling, the problems that present-day databases suffer from (inflexibility, fragility, etc.) will disappear. This claim has not yet been substantiated, although recently one system house, Logica, has developed a relational database package for use on a wide range of computers. Also IBM has, for several years, been developing a relational model which they call System R.

The concept of distributed databases is another related area of information storage. The distribution of information across a number of identical or similar machines can achieve the following benefits:

- Improved response times, because the multiple sites share the workload.
- Reduced operating costs, because several distributed machines can be less expensive than one central mainframe.
- Reduced costs to users, because they will incur only local call rates.

Distributing a database can, however, produce disadvantages. The major drawback comes from the problems involved in preserving the consistency of the database across all the separate machines. The problems are not insurmountable, but they do require a great deal of special consideration at the time when the system is being

designed. Often, system designers, either through lack of time or bad workmanship, build a database system first, and then try later to resolve the problems of the distributed aspects. When the system goes live, this approach inevitably produces difficulties.

Where the information on the database is of a regional nature, it makes sense to implement a distributed database, since information pertaining to a particular region need be stored only locally. The British Post Office viewdata service, Prestel, referred to on page 24, has adopted a distributed approach with its databases.

3. The retrieval of information

The system of retrieving information from a database must satisfy the following criteria:

- The command language to enable the user to retrieve the information must be easy to learn and easy to use.

- The user should be able to expect that he will not have to wait more than a few seconds for a response to his enquiry.

The users of on-line information retrieval systems can be classified as either public users or professional users. Increasingly, the public user has available to him information such as timetables, directories, and library catalogues that he can access quickly and easily through a computer terminal. For example, Kings Cross Station in London now has a computer terminal that provides timetable information.

The professional user (e.g. the researcher, the scientist, or the solicitor) on the other hand requires more facilities than the public user. An essential requirement is that the user should be able to search a large database on strings of text.

An important feature of an information retrieval system is the length of the time interval between the time of making the search and the time of receiving the hard copy.

Info-line Ltd., referred to on page 18, will be providing a direct ordering system as a feature of their system. Where the database contains numeric data, the retrieving of the data is often only a small part of the total task. Commonly, nowadays, users of information in the area of economics manipulate retrieved data through the use of extensive language facilities. Forecasting and modelling are important applications.

The Apple Corporation, who produce microcomputers, have developed some cassette-based software that is capable of accessing the Dow Jones database, retrieving 22 stock quotes, and storing the data on a floppy disc. Other software on the microcomputer enables the user to perform several types of analyses of the information.

The concept of relational databases was discussed on page 36. IBM have not only carried out extensive research with their System R. To complement it, they have also researched extensively into various languages which could be used to access information. The language which has reached the most advanced level of development is SEQUEL 2 (Sequential Query Language), which can be used interactively or as embedded verbs in PL/1. When used interactively it allows different users varying levels of facilities. At one end of the spectrum, the user can use the language after only a short training course. At the other end of the spectrum, the language represents a very powerful data manipulation facility. When they were developing the language, IBM took considerable care to ensure that it would be easy to use.

Although IBM have stated that System R is only a research project, and is not intended to produce a marketable product, it is reasonable to assume that the new range of mainframe computers that IBM are expected to announce shortly will provide some form of

relational database management system. The recently announced System 38 is reported to offer a relational database management system.

The majority of today's on-line information retrieval systems are accessed at a data rate of 300 bits per second. The introduction of facilities to access data at 1200 bits per second will give users two opportunities. They will either be able to perform the same searches as they do now and take only 25% of the time (and so achieve substantial financial benefits), or they will indulge in browsing, and so will spend the same amount of time and money as now but will get better value for it.

The availability of very high speed data rates in the future could lead to a situation where the intelligent terminal or microcomputer could be used. By retrieving large amounts of data very quickly and cheaply, and retaining this information on the local storage of the intelligent terminal or microcomputer (on floppy discs, cassettes, video discs, etc.), the user could then examine the information at his leisure by incurring only the costs of running an in-house intelligent terminal, or microcomputer.

It would be a natural development from existing intelligent terminals to produce software — for running on an intelligent terminal (microcomputer) system — that was capable of retrieving and manipulating data from an information retrieval system. This software would enable information retrieval systems to utilise future high-speed data links.

The concept of the multifunction workstation referred to in Report Series No. 4, Trends in Office Automation Technologies, where one terminal can act as an information retrieval terminal, a word processing system, an electronic mail transceiver and so on, is very important.

4. The output of information

Information in itself has no value: it becomes of value only when it is used. It is, therefore, important that the information that each different user requires should be tailored to suit his particular requirements.

As mentioned on page 7, searches on bibliographic databases normally result in the production of secondary information, such as references to publications or abstracts from publications. In some situations the abstract will satisfy the needs of the user, and can be output directly from the database either on-line or off-line. In the majority of situations, however, the usable information is contained in the publication itself (the primary information). The next stage in the information retrieval process, therefore, is that of obtaining the hard copy.

Present-day hard copy is usually provided on paper, but microform is increasingly being used. The ideal situation would be to hold all the primary information on the computer, and then to print on-line or off-line as required.

The majority of today's printers suffer from the limitation that they can print only the characters and symbols that have been built into the print mechanism, although some printers, of course, have interchangeable character sets. In the future, when databases hold multilingual text and graphics, the systems will require hard copy printing devices that have much greater capabilities than those possessed by current products.

A printer such as the ink-jet printer produces high-quality output. It does this by creating each individual character or symbol from tiny ink droplets which, under the control of software, are output to form the desired symbol or character. A high-speed ink-jet printer that had capabilities to print not only characters, but also high-quality graphics, would be well-suited to the task of handling the output requirements of the likely future information retrieval systems.

A product which has already been available for several years, but which is only now becoming widely accepted by users, is computer output microfilm (COM). Early COM devices were large complex pieces of equipment, and quite often these had wet photographic developing units which required substantial maintenance. Currently available equipment is much cleaner to handle, and it can be treated as simply another computer peripheral device. Clearly, the COM devices have considerable potential for being used as output devices with large information databases. It is interesting to note that neither of the market leaders in COM, Datagraphix Inc. or Eastman Kodak Inc., are computer manufacturers.

Both the ink-jet printer and the COM device are expensive pieces of equipment. Consequently, neither kind of device is likely to become part of an in-house system, except where there is a very high demand for information. The in-house searcher requires a low-cost product, and for him the facsimile device (fax) meets this criterion. Present-day fax is used for the exchange of information between persons. There is no reason, however, why information retrieval systems should not be the suppliers of the information.

Current fax equipment is classified by the CCITT as follows:

- Class 1 equipment — The transmission of an A4 page takes four to six minutes (analogue).
- Class 2 equipment — The transmission of an A4 page takes two to three minutes (analogue).
- Class 3 equipment — The transmission of an A4 page takes less than one minute (digital).

Very many Class 1 devices are already in use, and currently many of these are being replaced by Class 2 devices. Class 3 devices are also now beginning to be installed. A Class 3 device that is capable of receiving output from an information retrieval system represents a potentially very useful tool for the searcher.

As mentioned on page 23, the Japanese electronics company, Matsushita Electrical Co. Ltd., have developed a colour fax receiver which can be situated below a domestic TV receiver housed in the same unit. The combination of this type of equipment together with viewdata type systems could add a further dimension to the market for information.

What has been discussed above refers only to hard copy output devices. Soft copy devices, such as video displays, are not well suited to the task of retrieving bibliographic information (except for browsing purposes) because of the transient form in which the information is displayed. By contrast, when numeric data is being manipulated, the video display is essential. Where, in addition, the numeric data is re-interpreted into graphical representation, the user requires a high-quality video display unit that has facilities to produce, on request, a print-out of a screenful of data. Equipment that can be used in this way is currently available from companies like Tektronix Ltd. Over the next decade plasma displays are increasingly likely to replace the present cathode ray tube displays. (A plasma display is a flat display made up from a matrix of gas tubes.)

One product that is not yet available, but which embraces the benefits of both hard copy and soft copy, is the video disc. A guide to the commercial availability of the video disc was given as two years by Philips of Holland when they revealed details of their 1000 Mb laser disc recently. The video disc is well suited to replace some forms of paper publishing, since it combines the permanence of hard copy with the browsing capabilities of a video display. It also has two other advantages. First, because it is read optically, scratches and dust will not impair the legibility of the stored information, and second, it is very inexpensive. An impression of the possible cost of a video disc was given by

Martha Williams (of the Information Retrieval Research Laboratory in the US) when she spoke at the On-Line Conference in London during December 1977. She stated that the cost of a video disc holding one paperback book, which could be read from a hand-held viewer, was 10 cents. The concept of primary data being held on a computer and then being spooled out to video disc for distribution could, on the one hand, present a great opportunity and, on the other, pose a considerable threat to all forms of publishing.

5. The transmission of information

One of the major reasons behind the growth in the use of on-line databases has been the developments in transmission facilities. The transmission facilities do not only represent the umbilical cord between the user and the on-line information database; they also allow remote source data to be captured, and they provide the link to maintain distributed databases.

Unlike the other components of an information retrieval system (i.e. the capture, storage, retrieval and output of information), the future of transmission is not simply a matter of how quickly software and hardware engineers can design and produce products using the latest technology. The technical considerations must be balanced against the political considerations and, since all the PTTs in Europe have a monopoly, the political influences are significant.

A simple example to illustrate the situation is that of satellite communications. At the On-Line Revolution Seminar in Paris in July 1978, Alex Tomberg, the Chairman of EUSIDIC, urged that the possibility of a European Satellite Information Service be investigated. The figures shown in Figure 11 support his argument that satellite communication costs would be thirty times cheaper than the cost of present-day copper wire lines, and satellite communications certainly seem to warrant further study. There is little chance, however, that it would be possible to convince the European PTTs that such a project should be considered since the existing copper wires represent an enormous investment that is capable of earning revenue for several years yet.

Network	Cost per 1000 characters
ESANET	$0.92
TYMNET in N. America	$0.32
TELENET in N. America	$0.20
EURONET (proposal)	$0.16
TRANSPAC (proposal)	$0.068
EUSIS (European Satellite Information Service – proposal)	$0.00023

The above figures are calculated on the basis that searching at 300 bits per second produces an hourly throughput of 2,500 characters.

Source: Drawn from a presentation 'Networks in Europe', given by Alex Tomberg at the On-Line Revolution Seminar in Paris, July 1978.

FIGURE 11

In the future, as more countries develop national networks with facilities to connect with the EURONET network, access to EURONET will become easier.

As stated on page 15, the EEC have stated that they will consider applications from any non-EEC countries that wish to join EURONET. It is reasonable to assume, therefore, that the EURONET network will grow steadily during the next five years and, in so doing, will allow users in many countries to have access to more and more databases.

One significant factor that could affect the growth of EURONET is the tariff. The present tariff is geared to information traffic (i.e. small packets of information). However, information traffic constitutes no more than 2% of all data traffic. Bulk data traffic makes up the remainder, and it requires a tariff structure based on much larger packets. The way in which EURONET develops could have a significant effect on the future tariff structure for information traffic. If EURONET is opened to bulk data traffic, the outcome is likely to be a tariff structure which will increase the communication charges for information retrieval. On the other hand, if information traffic is actively encouraged, this could result in a decrease in communication charges for information retrieval. The latter seems to represent the more likely to produce future revenue for the PTTs.

C The Growth of On-line Information Services in the Future

As indicated on page 15, the future of European information retrieval services is closely linked with the future developments in EURONET. Those responsible for EURONET have been active in producing growth estimates for EURONET.

H. Ungerer, of the EEC, gave a paper at the On-Line Conference in London in December 1977 in which he suggested that EURONET will gain between 50% and 60% of the total information traffic for Europe. Based on an annual growth rate of 60% for the whole of Europe, EURONET could expect an annual traffic of around 1,100,000 searches by the end of 1983. This figure is currently being used by the EURONET marketing department as a marketing target. The extrapolated figure for the whole of Europe (assuming a growth rate of 60%) is over 2,000,000 searches per annum by the end of 1983. The thinking behind such estimates, according to Ungerer, is that:

- "With EURONET a new level of comfort of data service will be introduced into inter-EEC on-line searching".
- "Growth of this order has already been demonstrated in the US on a telecommunications infrastructure comparable to that of the future EURONET".

Clearly, there is no way of knowing at the moment whether or not these estimates will prove to be correct.

One significant factor in Europe which did not affect the early experience in the US is the existence of viewdata and similar systems that were referred to on pages 24 to 27. As was said there, when Prestel commences operation as a public service in 1979, it will introduce on-line information retrieval to a market (i.e. the business sector) that is still, in the main, ignorant about on-line information retrieval systems.

Prestel could well be the catalyst for the spread of on-line information retrieval services in Europe, and EURONET itself is certainly stimulating interest in on-line information retrieval systems. Provided that the tariffs can be kept at a realistic level, the evidence that is available today certainly indicates that the EURONET expectations are not unrealistic.

On the other hand, there are question marks to be put against these various components:

- Is Prestel sound enough from a technical point of view and will it provide the kind of information its potential users are likely to require?

- Will the European database vendors be able to compete successfully with the US database vendors?

The latter point is crucial in deciding whether there is an assured future for European information retrieval.

We believe that it is reasonable to expect that, by the end of 1983:

- Prestel will be an established information service in the UK, with a customer base of, perhaps, one million.

- Systems based on viewdata will have been introduced as a public service in several European countries.

- EURONET will be an established on-line information network providing on-line retrieval information facilities to users in both the EEC and the rest of Europe.

VII. THE ROLE AND RECOMMENDED ACTION FOR THE MANAGEMENT SERVICES DIVISION OF AN ORGANISATION WHICH IS USING OR ABOUT TO USE ON-LINE INFORMATION RETRIEVAL SYSTEMS

A Introduction

The introduction to this report emphasised the value of information to all businesses. All businesses make nearly all, if not all, of their important policy decisions without having all the up-to-date information they need for this purpose.

Mechanisms, in the form of information databases and their associated retrieval systems, are now becoming, or are about to become, available. They should permit businesses to have controlled and systematic access to information that is not only of a higher quality, but is also more relevant to the business needs, more accurate, more timely, and more comprehensive than is possible with present information systems.

This Section identifies the role that management services should play in ensuring that an organisation has access to, and exploits, this higher-quality information. In particular, it examines those steps that need to be taken to ensure that the organisation takes advantage of these new opportunities, both now and also in the next few years, as the available range of information continues to grow.

B The Role of Management Services

Management services departments have a threefold role in this field:

1. They have a responsibility for ensuring that available information systems (of which on-line information services are an example) are exploited to the full benefit of their organisations. They are therefore responsible for examining the scope for, and the feasibility of, using these services.

2. They have an essential part to play in the planning and implementation of each particular on-line information service because they are a source of expertise on the selection and installation of equipment.

3. They have an important role to play in co-ordinating all the various systems that are in use within their organisations, bearing in mind that an on-line information system will not be the only system in any organisation. They also have a role to play in ensuring that the overall systems plan in their organisations is a coherent one — and the implications of this are discussed on page 48.

   Management services also have a minor role to play as users of information services. The NCC database, THOR, is one example of an information service which is primarily directed at management services departments.

   Other departments also have a role to play in the use of on-line information services. Two examples are end users — who have an obvious role — and library or information units. The role of the library or information unit will vary from organisation to organisation depending on its responsibilities, previous experience and status.

C The Activities Associated with Exploiting On-line Information Services

As has already been mentioned, other parts of the organisation – the library, for example – have a part to play in ensuring that information services are used to the full benefit of the organisation. The department or unit that is in the best position to be responsible for information services will, of course, vary from organisation to organisation. This Section, therefore, is not specific about the respective parts that management services departments and information units should play. It concentrates on the activities which need to be undertaken.

As indicated earlier, the market for information services is emerging, and so much of the present activity in any organisation should be concerned with preparing for the future. Part of the preparatory stage involves identifying the responsibilities for the following:

- Monitoring the market place.
- Developing an awareness of the services among possible users in the organisation.
- Establishing procedures for evaluating users' requirements and available services.
- Selecting terminal equipment.
- Assisting in the initial use of the services.
- Ensuring that the requirements for accessing information systems are taken into account in the overall systems plan.

Each of these aspects is discussed below:

1. **Monitoring the market place**
   In order to ensure that users make the best possible use of information services it is important to monitor the market place for those new developments which could affect the organisation's users.

   On page 6 we suggested several organisations and publications that could help in this activity. Three ways of monitoring the market are:

   - Regular reading of those journals which contain articles and information concerning on-line information services.
   - Becoming members of those associations and organisations which have extensive involvement in the information services field.
   - Attending relevant seminars that cover aspects of information services.

2. **Developing an awareness of the services**
   In any organisation, the impetus to use information services needs to come from the end user. This can only happen if the potential users are fully aware of what is available, and also are able to assess the potential usefulness of higher-quality information. The vendors of information services do endeavour to sell those services to the end users in organisations and, in their selling activity, provide end users with a certain amount of knowledge of what is available. However, it may well be both prudent and sensible for organisations to set up a mechanism which ensures that information on the availability of services is circulated in a systematic manner to all concerned.

3. **Establishing procedures for evaluating requirements and services**
   Most organisations do not have procedures or methodologies that cover the evaluation of information services, either to determine the needs of the organisation or to determine the

appropriateness of a particular service.

The final responsibility for deciding on the use of a particular service may be taken centrally by (say) the librarian, or may be taken by the end user of the information. No matter who takes the decision there is an obvious need for all decisions to be co-ordinated centrally.

The central co-ordinator will be concerned with ensuring that the collective experience of the organisation is available to users. This experience is likely to encompass:

- The true costs of using an on-line information service.
- The likely capabilities and limitations of these services.
- The best sources of supply.
- The scope for negotiating contacts (including obtaining quantity discounts).
- The selection of suitable terminal equipment.

Standards are required for describing services and for evaluating them. Figures 12 and 13 are examples of checklists developed in the US for these purposes.

Evaluation exercises will be concerned with the cost/benefit of using a service. It will usually be easier to identify the costs than to identify the benefits.

The benefits will fall into two areas:

- Better information, leading to better decisions. The benefits in this area are extremely difficult to quantify.
- Increased efficiency. Productivity gains should result from the use of these services, since it should be possible to acquire higher quality information with less effort. This is an important opportunity since, at present, managers and professionals have few ways of obtaining direct assistance from information systems.

4. Selecting terminal equipment

Usually in selecting terminal equipment, there are not any considerations which are particular to the use of on-line information services. It is, nevertheless, sensible to make use of any experience and expertise that management services staff have in selecting terminal equipment.

In selecting terminal equipment, the points covered in 6 below need to be taken into account.

5. Assisting in the initial use of information services

The new user will undoubtedly find himself torn between doing the search himself and using an intermediary searcher.

The user will require some assistance from management services not only when deciding on which option to use, but also when using the selected option. If the user performs his own searches, he will certainly require assistance both in handling the terminal and in formulating the search. If he opts for using an intermediary, he will require some guidance on how to make the best use of the searches – for example, by holding lengthy discussions with the searcher in order to establish exactly what he (the user) requires from the search.

A. Content (identification and contents)

1. Name: What is the name of the databank?

2. UOE
(a) What is the unit of enumeration (UOE)? Examples: corporation, security, commodity contract, economic time series, legal opinion, bibliographic abstract, and patient record.
(b) How are specific UOEs identified? Example: a company might be identified by its CUSIP issuer number.
(c) Who assigns the identifying codes?
(d) Do the codes ever change? If so, how often do they change, and how are the changes cross-indexed?

3. Coverage
(a) What is the basis for including a specific UOE in the databank? Example: all common stock on the New York Stock Exchange, all decisions of the US Supreme Court, all reviews in *Computing Reviews*.
(b) What data are carried for each IOE?
(c) What percentage of the total number of possible data items is "missing", and how are those items identified on the databank (i.e., where are the gaps and how important are they?)?
(d) What happens to a UOE that is dropped?

4. Classifications
(a) What generally recognized classifications are used on the databank? Examples: CUSIP security identifiers, stock exchange ticker symbols, ZIP codes, telephone area (DDD) codes, state abbreviations, Standard Metropolitan Statistical Area (SMSA) designations, Standard Industrial Classification (SIC) codes, Federal Reserve districts, federal stock numbers, Universal product codes, Social Security account numbers, West Publishing Company headnote codes, and so on.
(b) What classifications are used that may be useful elsewhere?
(c) Are there ways of connecting any generally recognized classifications and the classifications internal to the databank?

5. Plans: Are there any plans to change the contents of this databank?

B. Form (form in which the database exists and is available)

1. Format
(a) What is the primary medium that the databank is stored in?
(b) What is its physical record layout? (Questions (b) and (c) are for the benefit of the technical staff.)

2. Access
(a) In what form is the databank available to users? Examples: fixed format reports, tailored reports, interactive access, machine readable subsets, complete file, initial file plus updates, and so on.
(b) For each form, what is the medium, the file organization, and the record layout?

3. Restrictions
(a) What limits are placed on the user's rights to redistribute either the raw data or any form of the data processed from the databank?
(b) What special equipment must the user have to be able to use the databank?

4. Plans: Are there any plans to make any change in the form of this databank?

C. Logistics and economics (logistical and economic trade-offs)

1. Timing
(a) How frequently is the databank updated?
(b) How fresh are the data which are used for updating?
(c) How soon after updating is the databank available for distribution?
(d) Is there any time saving or penalty associated with using a republisher (intermediary) rather than a vendor?

2. Correctness
(a) What is the error rate expressed as a percentage of the total number of data items and as a percentage of the number of UOEs?
(b) How are errors detected and corrected?
(c) Are there any internal inconsistencies inherent in the databank because of, for example, changes in estimation methods, sampling procedures, or units of measurement?
(d) Were items restated or adjusted and, if so, why?

3. Cost: What is the *full* cost of the databank to a user? Include library fees, updating fees, time-sharing charges, surcharges, fees for frequent updates and early deliveries, per usage fees, package and quantity discounts, tie-ins with other products, incremental equipment costs, costs of training, and extra personnel.

4. Plans: What plans of the vendors or republishers would affect the logistics and economics of using this databank?

D. Interfaces (human and other)

1. Publishers, i.e., sellers
(a) For the publisher of this databank, what is the publisher's name, address(es), telephone numbers for information and time-shared access, and names of sales and technical representatives?
(b) For all republishers, what are answers to same questions asked in (a)?
(c) Who is to be called on in case of need for advice, technical problems with use of data, data error, late delivery, change of user needs, and other problems?
(d) What aids are available to the user? Examples: presales help, testing periods, postsales support, documentation, and software.

2. Users
(a) How will the databank be used?
(b) Who will use the databank, in what ways, and how extensively? Can any of the users furnish references?
(c) Does a users' association exist and, if so, where?

3. Other interfaces
(a) Aside from publishers and users, who sponsors, regulates, or otherwise influences the databank?
(b) Can this databank be used easily with any other databanks?
(c) What other databanks are available from these sellers (publishers and vendors) and intermediaries?
(d) How secure and how private are user proprietary data, software, and systems?

4. Plans: Are there any plans that might affect the human and other interfaces associated with this databank?

FIGURE 12 A DATABANK QUESTIONNAIRE – USED FOR PRODUCING A DESCRIPTION OF A SPECIFIC DATABANK

Source: Harvard Business Review Nov-Dec 1978

1. What is the application for which data is needed?

---

2. What data can (or must) be obtained from outside sources?

---

3. What attributes of an outside databank are ***essential*** to the success of the application, and what attributes are ***desirable?*** Examples:
   (a) Coverage (how much, how far back in time, and what items?)
   (b) Classifications (which ones?)
   (c) Format (what file organisation medium, and physical structure?)
   (d) Access (what form and method?)
   (e) Restrictions (how onerous?)
   (f) Timing (how soon, how often?)
   (g) Correctness (how "clean"?)
   (h) Cost (how much?)
   (i) Publishers, intermediaries (who, offering what help?)
   (j) Users (what groups might use it?)
   (k) Other (what compatibilities, how secure?)

---

4. What databanks are available that might be satisfactory? Is there enough benefit in adjusting our demands or in gathering the data ourselves to make such actions worthwhile?

---

5. For purposes of this application, what flaws does each databank have? Examples:
   (a) Inaccurate?
   (b) Inconsistent?
   (c) Incompatible?
   (d) Incomplete?
   (e) Unrepresentative?
   (f) Unwieldy?
   (g) Uneconomical?
   (h) Untimely?

FIGURE 13 A DATABANK APPLICATION EVALUATION – USED TO ASSESS HOW WELL THE DATABANK IS SUITED TO A PARTICULAR NEED

Source: Harvard Business Review Nov-Dec 1978

6. **Ensuring that the systems plans incorporate the requirements for accessing information services**
Properly devised, an organisation's systems plan takes account of a whole range of requirements for information systems, including the accessing of public information services. Terminal equipment is used for this purpose, but it is likely, in many cases, to be a low-activity use and, therefore, difficult to justify for this use alone. By careful planning it may, instead, be possible to use general purpose devices which are already being used for (and are mainly being justified by) other applications. An organisation, when it purchases general purpose devices in future, should bear in mind that they may be required later to access an on-line information service. This applies even though the organisation's present plans do not include the use of on-line information services.

D Recommended Action

This report has described the way in which on-line information services have already developed, and can be expected to develop in the near future. Also it has discussed the role management services should play in exploiting these services, and has set out the activities which an organisation needs to undertake. The specific action which organisations should take will vary both according to their previous experience in this field and also upon the extent to which they can exploit on-line information services. However, all organisations can review the use they already make of these services and they can assess the effectiveness with which they exploit those services. They can also consider to what extent and in which way they would benefit from using those services in the future.

We believe that management services will usually be best placed to undertake this review since it is they who have the combination of skills and experience which is needed. Obviously other departments — such as the library — also need to be involved. Such reviews can be conducted as individual projects with the normal controls and reports which are produced by feasibility studies and system audits.

VIII. CONCLUSION

At the beginning of this report we set out to answer the three following questions:

1. Is on-line information retrieval a subject that is sufficiently important to warrant organisations devoting attention to it now?

2. If so, what action should management services managers be taking, both now and in the longer term, to prepare for on-line information retrieval?

3. When is on-line information retrieval likely to have a significant impact on business organisations and on their day-to-day operations?

It is clear that on-line information retrieval services will become increasingly important in the next few years as more services become available both to the specialist and to the more general user. The attention which management services need to pay to the subject today is concerned with evaluating existing services in conjunction with users who might make beneficial use of them and with establishing a framework within which to monitor the market and co-ordinate the use of services in the future.

In terms of what action management services should take, the report has outlined measures which will assist organisations currently using these services and organisations which are not yet using the services, in reviewing their requirements, and the way in which those requirements are met.

The third question remains a difficult one to answer since there are a number of relevant commercial and technical issues. The impact of viewdata systems and of EURONET and the marketing decisions of those promoting the databases cannot be accurately assessed at this time.

However, it is clear that almost all developments point in the direction of an increased usage of on-line information services over the next few years which may well lead to them being commonplace within the next five years.

BIBLIOGRAPHY

Tony Gunton — Terminal Compatibility (Butler Cox Foundation Report Series No. 3)

Tony Gunton/Roger Camrass — Public Data Services (Butler Cox Foundation Report Series No. 7)

John Kinnear — Trends in Office Automation Technologies (Butler Cox Foundation Report Series No. 4)

Philip Williams — Report on International On-line Information Systems (prepared for Butler Cox & Partners July 1978)

Ariane Iljon — Scientific and technical databases in a multi-lingual society (On-line Review June 1977)

Alex Tomberg
1. On-line Services in Europe (On-line Review Volume No. 3 1977)
2. Networks in Europe (presentation at the On-line Revolution Seminar ICSU AB July 1978)

Mike Lea — Associative Processing of Non-Numerical Information (Technical Memorandum No. C/SR/060 Brunel University)

Martha Williams
1. A paper entitled "On-line retrieval — today and tomorrow" given at the On-line Conference December 1977
2. Database and On-line Statistics — 1977 (Bulletin of ASIS December 1977 Vol 4 No 2)

T. Flynn and A.E. Wild — A paper entitled "Integrated on-line information systems in a large multinational company" given at the On-line Conference December 1977

Haines Gaffner — The LINK study — Strategies in the on-line database market place

Joel W. Darrow and James R. Belilove — The growth of databank storing — Harvard Business Review Nov-Dec 1978

Garth W. P. Davies — The EURONET Project — Paper given at 1978 NFAIS annual conference

Abstract

Report Series
No 10

Public On-line Information Retrieval Services

by Peter Appleby
November 1978

ABSTRACT

On-line information services are now appearing in Europe, and these offer all organisations new ways to obtain information from outside. This external information is needed by managers and professional staff in almost all parts of an organisation.

These information services, which enable an end user to obtain information which is relevant to his needs, represent one way of overcoming the problem of an ever-increasing supply of published material.

The report examines the development of information services in the US, Europe and Japan and discusses the directions in which those services can be expected to expand in the next few years through a number of relevant technological advances.

The report offers advice to organisations on the action they need to take to ensure that they make full use of the currently-available services, and that they equip themselves to exploit new services as they reach the market-place.

The Butler Cox Foundation

The Butler Cox Foundation is a research group which examines major developments in its field — computers, telecommunications, and office automation — on behalf of subscribing members. It provides a set of 'eyes and ears' on the world for the systems departments of some of Europe's largest concerns.

The Foundation collects its information in Europe and the US, where it has offices through its associated company. It transmits its findings to members in three main ways:

- As regular *written reports,* giving detailed findings and substantiating evidence.

- Through *management conferences,* stressing the policy implications of the subjects studied for management services directors and their senior colleagues.

- Through *professional and technical seminars,* where the members' own specialist managers and technicians can meet with the Foundation research teams to review their findings in depth.

The Foundation is controlled by a Management Board upon which the members are represented. Its responsibilities include the selection of topics for research, and approval of the Foundation's annual report and accounts, showing how the subscribed research funds have been employed.