From Babel to EDIL: the evolution of a standard for document delivery

Andrew Braid <Andrew.Braid@bl.uk>

Abstract

There is a great deal of interest in electronic document delivery at the present time. This paper examines electronic document delivery and one of the many problems which it raises, that of standardisation. The paper details the development and proposals of one standard, GEDI, and gives details of a pilot demonstration of the standard in four European countries.

I. Introduction

Electronic document delivery is a "hot" topic at the moment. Readers are becoming aware of the many advantages provided by the desk top delivery of documents. Libraries are looking at the possibilities that electronic document delivery provides for changing from a holdings (or "just in case") to an access (or "just in time") policy. Document supply organisations are looking at ways of providing faster services to readers using electronic document delivery. Other groups, especially publishers, are concerned about the economic effects, particularly the possible loss of revenue.

As is often the case with new developments, many of these activities are taking place independently. Little work has been done in the area of standardisation of electronic document delivery; the chances of an electronic Babel developing seem very high. This paper outlines one piece of work which has been done and also gives details of a pilot implementation involving users in four European countries.

II. The Information Chain

Electronic document delivery is part of the chain of information provision. For electronic document delivery, the assumption has to be made that the required information is contained in a document which is not available locally. A simple chain in the requesting of a document from a remote source might be as follows, (from the requester's view):

- identification of required information (or document)
- identification of the source of supply
- ordering the document
- receiving the document
- paying for the document

Some of these processes may be carried out by an intermediary such as a librarian.

The supplier has additional steps in the chain. The principal ones are:

- receipt of request
- processing the request
- copying the document
- posting the document.
- accounting/billing procedures

Electronic document delivery has tended to be used as a faster version of conventional document delivery, in which a photocopy of the requested article is posted to the requester. Little change is required in the steps outlined above if electronic document delivery is substituted for conventional delivery. The two main changes are that the receiver has to print the document (assuming that it will not be read on the screen) and the supplier has to scan the document instead of copying it.

III. Definition of electronic document delivery

It is necessary to clearly define electronic document delivery. At one level it is very simple; it is the electronic transmission of a document. Questions immediately arise, such as how and in what format will the document be transmitted. The majority of electronic document delivery falls into two categories. The first is the transmission of files in character encoded format. The other is facsimile transmission. The latter is now widespread, with possibly as many as 20 million machines world-wide. Both methods have strengths and weaknesses. Character encoded documents are much more compact and are capable of easy processing, searching, etc., but documents transmitted in this format can contain little non-textual material. Facsimile encoding can readily include any representation on a printed page, including colour. Its major disadvantages are the bulk of image files and the lack of searchability.

Recent advances in technology, especially network capacity and the processing power of desk top computers, has led to a blurring of the distinction between these two systems. Various

systems are now available which can describe the content of a document and others can describe the layout of a document in character encoded format. Most of these systems still have great difficulty with pictures and other non-text material. For this reason, libraries have tended to use facsimile transmission. This allows the receiver to have a true copy of the original and emulates the existing practice of document delivery by photocopying. This paper concentrates on the transmission of page images, although many of the processes described are capable of transmitting character encoded documents.

A general model for the interchange process is shown in Fig. 1. The main characteristics of the model can be described as follows:

- the interchange involves two parties, the supplier and the consumer;
- the supplier and the consumer are linked through a facility enabling the transfer of an electronic document from supplier to consumer;
- the transfer handles one document at a time.

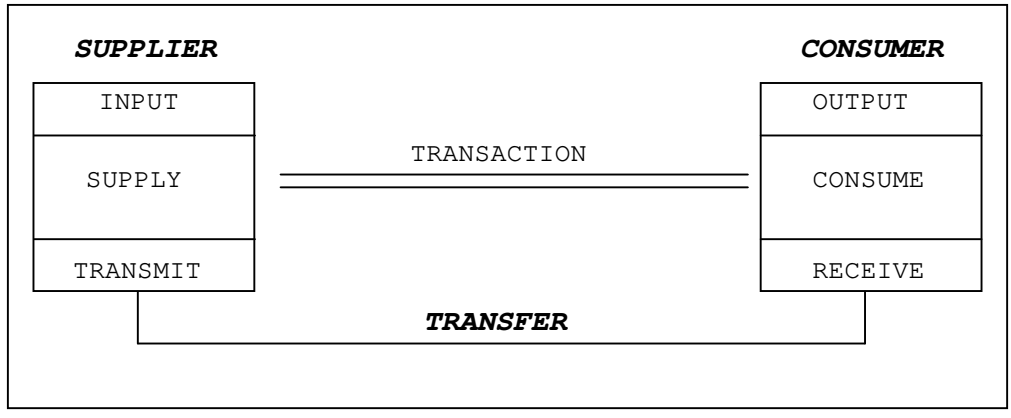

Fig. 1 General Model for Document Delivery Interchange

The complete cycle of interchange, starting with an electronic document residing at the supplier and terminating with the completed delivery of that document to the consumer is called an electronic document interchange transaction. It is important to note that the input and output functions as shown in Fig. 1 do not participate in the transaction. In practice, some form of input and output will be available. Input may be from hard-copy documents through scanning, from files of stored images or from electronically published documents; output can be take the form of electronic storage or printing. The implementation of some of these possibilities may be dependent on legal and copyright regulations - see section 13.

IV. The Development of a Standard

In an attempt to avoid an electronic Babel several libraries and other interested parties met in late 1990 to discuss the possibility of developing a standard for the electronic delivery of documents in image format. Those present included representatives from the following:

France:

The Ministere de l'Education Nationale Sous-Direction des Bibliotheques (MEN/SDB)
Questel
Telesystemes

Germany:

Universitatsbibliothek Hannover und Technische Informationsbibliothek (UB/TIB)

The Netherlands:

Pica - Centrum voor Bibliotheek-automatisering

United Kingdom:

The British Library Document Supply Centre (BLDSC)

United States of America:

The Online Computer Library Centre (OCLC)
The Research Libraries Group (RLG)

The work was partially funded by the European Commission (DG XIIIB)

All agreed to the development of a standard for electronic document delivery. The group was called GEDI - the Group on Electronic Document Interchange. Several meetings followed, and in surprisingly short time, agreement was reached and a Recommendation was published in October 1991 [1].

V. The GEDI Recommendation

The Group made several decisions before finalising its Recommendation. It was agreed that common, international agreement on all the functional and technical details of electronic document delivery. The GEDI Recommendation is, therefore, based on the concept of domains. A domain is defined as a group of one or more suppliers and one or more consumer capable of engaging in electronic document delivery transactions (Fig. 2). This concept means that solutions can be developed under private responsibilities in private domains that are more or less independent of the solution used in the common, international, GEDI domain. Common

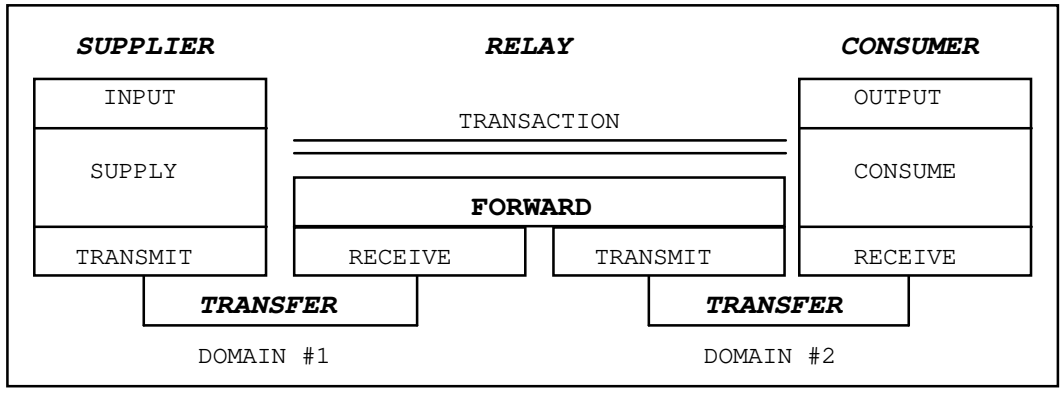

Fig. 2 Interchange across two domains

GEDI should support existing standards, protocols and profiles as much as possible. It was recognised that electronic document delivery would take place between a wide variety of organisations, some of which would already have electronic document delivery systems in place. It was, therefore, agreed that GEDI would concentrate on the document interchange format and a description of the interchange mechanism.

It was recognised that it would not be practical (certainly within the short time scale) to reach a agreement is required only within the GEDI domain; GEDI agreements may or may not be implemented in private domains. One important part of the GEDI Recommendation is to describe the boundary between the private domains and the GEDI domain, especially the relay functions for such boundaries (Fig. 3).

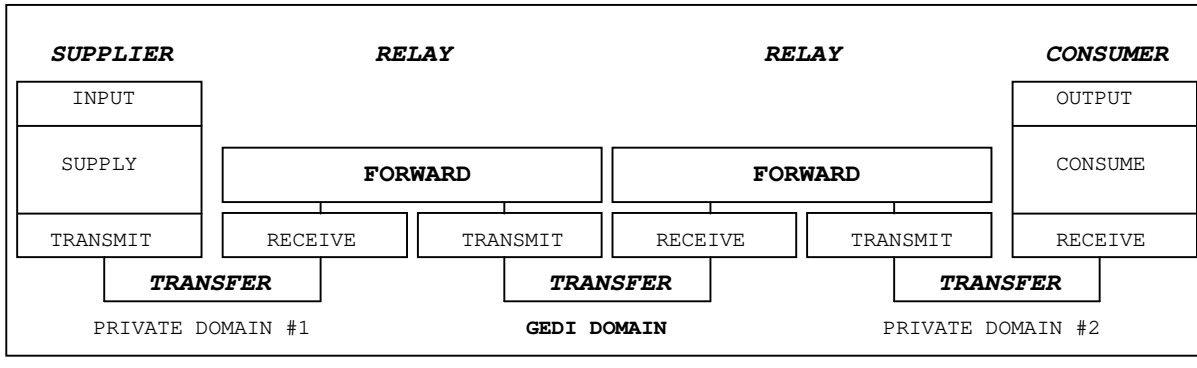

Fig. 3 Communication through the GEDI Domain

The GEDI Recommendation described how a document should be transmitted as well as the format of the document. In brief the recommendation was as follows:

Resolution:	300 dpi
Compression:	CCITT Group 4
File interchange format:	TIFFv5
Document transfer mechanism	FTAM
Network protocol:	X.25

The document interchange format structure consists of two parts; the cover information and the document image. The cover information, or GEDI header, describes, in four classes, various attributes of the document - see Fig. 4

Because most of the members of GEDI were closely involved in the development of two library OSI application protocols, Search and Retrieve [2] and Inter Library Loan [3], many of the fields used in the GEDI classes are closely related to those protocols. This greatly facilitates the integration of the searching, ordering and delivery processes. It should be possible to use several elements of an electronic request to generate the GEDI header.

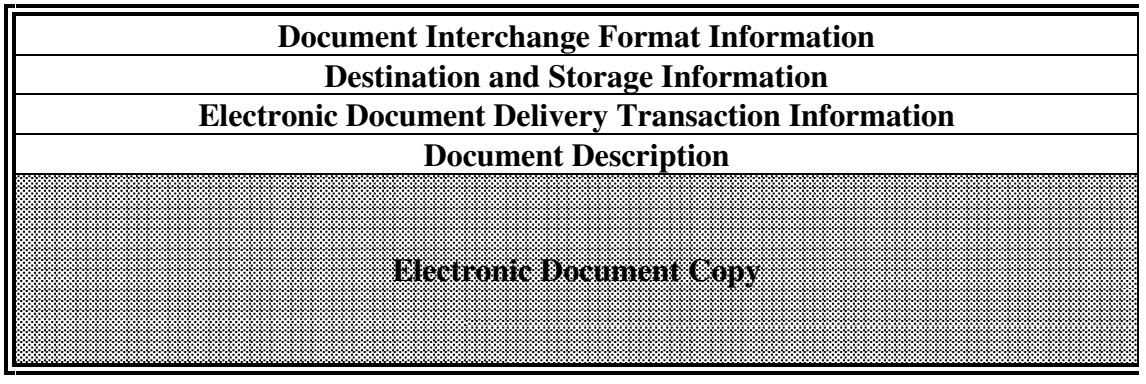

Fig. 4 GEDI Document Interchange Format

VI. Practical Implementation

Although all the members had a mutual interest in the electronic exchange of the full text of documents, none of them had any immediate plans to implement the GEDI Recommendation. In fact, the only members which had an operational system were RLG, with its Ariel electronic document delivery workstation [4], and MEN/SDB with the FOUDRE system [5]. However, all the GEDI members did have medium term plans to provide electronic document delivery services and all were obviously committed to GEDI. The European members, with the addition of a further partner from France, (the Institut de l'Information Scientifique et Technique [INIST]), agreed to make a submission under the CEC Library Programme to test the Recommendations. A proposal was submitted in December 1991 for a system, using the GEDI Recommendation, to link the proposed electronic document delivery systems of Pica, TIB, MEN, INIST and BLDSC with Telesystemes as Project Manager. The original partners were joined by FCCN in Portugal who are responsible for dissemination of information about the project. The consortium bid was successful and in January 1993 the EDIL Project (Electronic Document Interchange between Libraries) was born. It is an ambitious project. The total cost is 2 million ECU (about $2 million) over a 30 month period.

VII. Background of the Partners

The various partners all have different methods of working as far as document delivery is concerned and all have different technical infrastructures. The one thing which the partners do have in common is a wish to improve the existing methods of document delivery.

VII.A. Organisational

In France, there are two separate systems for document supply. The CADIST system of subject specialisation is operated by MEN/SDB. Separate university libraries have collections of journals and periodicals in differing subject areas. These libraries make these collections available to all other university libraries in France. INIST works on a totally different principal. It has a central collection of journals and periodicals in the fields of science, technology and medicine which it makes available to users from its building in Nancy in eastern France.

In Germany, there is also a system of subject specialisation, but in this case it is for all libraries, not just academic libraries. TIB holds the collection for all scientific, technical and engineering subjects. When a local library does not hold an item in these subject areas and it is not held

elsewhere in the local region, it forwards the request to TIB.

In the Netherlands, Pica operates a periodical union catalogue system. The periodical holdings of all Pica members are held on a central, automated catalogue. A highly sophisticated ILL requesting system is also based around this automated catalogue. Requests are sent to Pica, where an automated system identifies potential suppliers and forwards requests.

In the UK there is a hybrid system. The majority of requests are satisfied from the central stock held by BLDSC at Boston Spa, but requests for items not held in the central collection are forwarded to other libraries either in the UK or abroad. Unlike most of the other partners, BLDSC also receives requests from international customers. This means that BLDSC has, for a long time, looked at ways of reducing the time for the delivery of documents.

VII.B. Technical

The only partner with any practical experience of an electronic document delivery system was MEN/SDB who had implemented the FOUDRE system. This system had been developed by Telesystemes and had been implemented in 6 French universities. In the FOUDRE project documents are scanned and compressed using the same protocols as GEDI but the file transfer is STUTEL and Numeris, the French ISDN network, is used for transmission. Articles are stored on an optical disk after scanning in case of any future use. INIST has installed a large scale optical storage system for 1,800 journal titles, but little of the output is transmitted to customers in electronic format.

In the UK, BLDSC is using the ADONIS system, which stores some 450 journals on CD-ROM, but again there is no link to electronic document delivery. BLDSC is formulating plans for electronic document delivery. It has carried out several trials with a variety of systems and networks [6]. At the time the EDIL Project was launched, BLDSC was planning a system for electronic document delivery over JANET, the UK Academic network. This system has implemented the GEDI Recommendations but uses electronic mail, X.400, as the transport mechanism [7].

In the Netherlands, Pica is planning a system called RAPDOC [8]. This plans to make information about individual articles available on the existing database of periodical titles and assigning supply of specified periodical titles to designated libraries to guarantee fulfilment. Pica plans to use a modified version of the RLG Ariel workstation. In its original form this was not conformant to the GEDI Recommendation; it has no header information and it uses the CCITT group 3 compression algorithm (T.4) and TFTP. The modified version is planned to be much more GEDI conformant - the major divergence being the use of FTP and not FTAM.

In Germany, little progress had been made with electronic document delivery. In 1993 the regional library system of Lower Saxony, in which TIB is situated, implemented the Pica system and thus can use the RAPDOC system.

VIII. The EDIL Project

The EDIL Project began in January 1993. The Project is implementing a pilot electronic document delivery system between the private domains of the five partners. In spite of the delay since the GEDI Recommendations were issued, EDIL will be the first practical test of the Recommendations. The major deliverable of the Project will be the implementation of four interlinked relays, one in each country of the partners. A secondary aim is the dissemination of information about the Relay to all interested organisations both inside and outside Europe.

The Project has 6 main stages. These are as follows:

- Requirements
- Specifications
- Development
- Integration
- Installation and Testing
- Use and Evaluation

Different partners are responsible for each stage. Requirements were split into User Requirements, for which BLDSC was responsible, and Technical Requirements, which were co-ordinated by Pica. Both were completed in July 1993. The Specifications were prepared by INIST in March 1994. The Development stage is to be carried out by Telesystemes, who are also Project Leader. Each Partner is responsible for Integration with its national system and for Installation and Testing. Use of the system will be co-ordinated by MEN/SDB and BLDSC is responsible for Evaluation.

Each country will have a single Relay. As well as the main suppliers in France, Germany and the

UK, there will up to 5 test site libraries in those countries. These test sites will act mainly as requesters although a small number will also act as suppliers. In the Netherlands there will be about 5 test sites which will act as both requester and supplier.

The Relays will be connected using various networks. In the UK, the SuperJANET network will probably be used. SURFNET will be used in the Netherlands and WIN in Germany. The position in France is not clear at present. The France Telecom ISDN network, Numeris, was used for the FOUDRE system and not RENATER. The latter has been upgraded and may be used for EDIL. The Relays will be interconnected using either Europanet or Ebone

The analysis of User Requirements of the system was developed from a combination of questionnaires and meetings of potential test site libraries. In some cases, e.g. some French university libraries, previous experience of using an electronic document delivery system meant that it was fairly easy to specify details of the level of service required and how the proposed trial system would be integrated with existing procedures. In other cases, it proved quite difficult to obtain this information. This was because electronic document delivery allows the possibility of delivery directly to the desk of the end-user. There was a great deal of difference in the needs of end-users in different market sectors. These differences were even more apparent in the price that end-users were prepared to pay.

The Technical Requirements of the Relay were defined fairly easily. However, a great deal of effort went into the requirements for the document ordering system. It had originally been intended not to cover this aspect, but during the preparation of the Technical Requirements, it became clear that certain tasks, especially those concerned with the preparation of the GEDI header page information, were inextricably linked to the requesting process. It was, therefore, necessary to co-ordinate the ordering process and integrate it with the delivery process.

As mentioned earlier, most of the Partners had been involved in the development of the OSI ILL protocol. Many of the fields in the GEDI Header are derived from the ILL protocol. It was agreed, therefore, that the use of these fields would simplify the production of the GEDI Header for the EDIL system. Only two of the Partners, Pica and MEN/SDB had implemented the OSI ILL protocol. It was agreed that it would not be necessary to implement the full OSI ILL Protocol as long as the required fields were included in requests.

VIII.A. Potential Traffic

One problem which was foreseen at an early stage was the low volume of documents delivered between the Partners. BLDSC is the only supplier with any appreciable business with the other countries. Each partner also charged different tariffs for documents supplied. Left to natural forces, there would be little traffic passing through the Relays. In this situation, it would not be possible to achieve the principal objective of the project, namely to prove the technical feasibility of the GEDI Relay. It is not a prime objective of the project to demonstrate the benefits of international electronic document delivery, although this is a secondary aim of many of the partners. The Technical Specifications state that each Relay should be capable of a throughput (i.e. either sending or receiving) of 100 documents per day. Natural traffic is estimated at no more than 10% of this total. A proposal has been made to route requests artificially, which would be real requests from the test site libraries, to other suppliers so that a copy of the requested document can be transmitted through the Relay. At the time of writing the exact details of this procedure have not been agreed by the Partners.

VIII.B. Technical Solution

The final technical design of the Relay had also not been finalised at the time of writing. The following description contains the latest information that is available and may change over the next few months as the project progresses.

The hardware platform for the Relay will be either an IBM RS 6000 or a Sun Sparcstation. As well as allowing the implementation of the full GEDI Recommendation, of FTAM over X.25, this solution will allow a variety of other options. These will include FTP over TCP/IP and X.400 over either X.25 or TCP/IP (RFC 1006). It is the responsibility of each partner to implement the national side of the EDIL Relay and each partner will adopt a different technical solution. In the UK documents will be transmitted from BLDSC to the Relay in GEDI format but under the X.400 protocol. This will be converted by the Relay to the FTAM protocol for forwarding to the other Relays.

In the reverse direction, documents received at the Relay under FTAM will be converted to X.400 and forwarded to the recipient. In this way, the

receiving of documents from either national or international sources will be transparent to the test site libraries.

VIII.C. The Future of EDIL

The EDIL Project is a large and complex one. It involves 7 partners from 5 European countries. The total time of the Project is estimated at 30 months and due to end in August 1995. At the time of writing (March 1994) the project is about 8 weeks behind schedule. The Relay should be available for testing during late Summer 1994 and the pilot experiments should take place in the first 6 months of 1995.

IX. Conclusion

EDIL is probably the first example of inter working differing electronic document delivery systems. It is an ambitious project and it is too soon to say whether it will succeed. At present the signs are that it will be successful at the technical level. It is far from clear if it will succeed in other ways. There are many obstacles to be overcome. One of the most important is the issue of copyright. As stated in the Introduction, publishers regard all forms of electronic document delivery with a great degree of uncertainty. They fear that if uncontrolled, electronic document delivery would be a severe threat to their revenue. Their argument is that if quick electronic access is readily available, libraries will cancel subscriptions in favour of individual article supply from a remote source.

There is no easy answer to this question. Many publishers are now experimenting with a variety of methods of making primary journals available in electronic format. Few of these experiments can be regarded as outstandingly successful. One of the main reasons for this is a reluctance to change on the part of users. This also applies to electronic document delivery systems. It was discovered as part of the User Requirements for EDIL that users are not yet prepared to read documents on screens. They would far prefer paper output. Until users wish to keep documents in entirely electronic format, it will be necessary to design systems which make a printed copy. EDIL has been designed with this need in mind. In a few years time this may change as a new generation becomes more computer literate.

X. References

[1] "Electronic Document Delivery: Towards Further Standardisation of International Interchange. Agreements of the Group on Electronic Document Interchange," .October 1991, Unpublished. Copies are available from the GEDI Secretariat c/o Pica, PO Box 876, 2300 AW Leiden, The Netherlands

[2] ISO 10162 Information and Documentation - Open Systems Interconnection (OSI) - Search and Retrieval Service Definition - 1991

ISO 10163 Information and Documentation - Open Systems Interconnection (OSI) - Search and Retrieval Application Protocol Specification - 1991

[3] ISO 10160 Information and Documentation - Open Systems Interconnection (OSI) - Interlibrary Loan Application Service Definition - 1991

ISO 10161 Information and Documentation - Open Systems Interconnection (OSI) - Interlibrary Loan Application Protocol Specification - 1991

[4] Mary E. Jackson, "Using Ariel, RLG's document transmission system to improve document delivery in the United States" Interlending and Document Supply, vol. 20 no2, pp49-52, April 1992

[5] Christine Deschamps, "FOUDRE - electronic access to documents in the French academic community" Interlending and Document Supply, vol. 19, no 4, pp127-130, October 1991

[6] J Andrew Braid, "Electronic document delivery: a reality at last?" ASLIB Proceedings, vol. 45 no 6, pp161-166, June 1993

[7] Bill Tuck and Ruth Moulton, "Using X.400 for document delivery" Computer Networks and ISDN Systems, vol. 26 no 3, pp291-296, 1993

[8] Look Costers and Sjoerd Koopman, "The Dutch RAPDOC project: from interlibrary loan to electronic document delivery," Interlending and Document Supply, vol. 21 no1, pp4-6, January 1993

Author Information

Andrew Braid joined the former National Lending Library for Science and Technology at Boston Spa in 1968 after graduating from the University of Birmingham. He has held a variety of posts including management of the literature stores and reprographic unit, the conservation unit, stock relocation and building planning.

In 1984 he moved to a newly created post of Service Development Officer. This involved examining the use of new technology for the services offered by BLDSC. Initially, this was in the area of electronic document storage and delivery. He was chairman of the ADONIS Document Supply Centre Liaison Group as well as specifying the ADONIS workstation. He represented the UK in the APOLLO project for satellite document delivery.

In 1986 he became Head of Service Development. He headed an internal team which introduced an electronic request processing system within BLDSC and was responsible for the marketing of BLDSC services.

Current projects include: the capture of journal contents pages, the electronic scanning of full text, automated matching and sorting of electronic requests, the use of facsimile image transmission by ISDN, academic networks and satellite and the integration of these processes. He is particularly interested in the organisational changes that the use of new technology will bring about and has written several articles on various aspects of the introduction of new technology to library services.

Currently, he is Chairman of the International Forum for Open Bibliographic Systems (IFOBS), an international group which is looking at the problems of implementing OSI protocols in libraries. He is also a member of the Group on Electronic Interchange (GEDI) and is the UK member of the EDIL Management Board.