AGARD-CP-544

AGARD-CP-544

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS 544

International High Speed Networks for Scientific and Technical Information

(Les Réseaux internationaux rapides
d'échange d'information scientifique et technique)

Papers presented at the Technical Information Panel Specialists' Meeting held at the National Library of Canada, Ottawa, Canada, 6th—7th October 1993.

NORTH ATLANTIC TREATY ORGANIZATION

Published February 1994

Distribution and Availability on Back Cover

AGARD-CP-544

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS 544

International High Speed Networks for Scientific and Technical Information

(Les Réseaux internationaux rapides
d'échange d'information scientifique et technique)

Papers presented at the Technical Information Panel
Specialists' Meeting held at the National Library of Canada,
Ottawa, Canada, 6th—7th October 1993.

North Atlantic Treaty Organization

Organisation du Traité de l'Atlantique Nord

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from material supplied by AGARD or the authors.

Published February 1994

ISBN 92-835-0736-3

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

Theme

Wide area electronic networks have been developing rapidly in recent years, and are being interlinked to form an international network. Information centres will have a tremendous opportunity to make use of these networks for the dissemination of STI in aerospace, defence and other fields. On the other hand, these networks also give rise to many new concerns and issues. This Specialists' Meeting of participants from NATO countries reviewed the current situation, assessed future possibilities and examined issues, such as standards, security, management, property rights/copyright, etc., from the point of view of the end user, the information provider, and the network manager.

Thème

Au cours des dernières années, les grands réseaux ont connu une expansion rapide. Aujourd'hui, en se reliant les uns aux autres, ils jettent les ponts d'un réseau international. Les centres d'information peuvent maintenant profiter de cet outil exceptionnel pour diffuser de l'IST sur l'aérospatiale, la défense et sur d'autres domaines. Toutefois, l'émergence de ces réseaux laisse poindre certaines préoccupations et questions. Cette réunion de spécialistes des pays de l'OTAN a permis de faire le survol de la situation, d'évaluer le potentiel d'avenir et d'examiner des questions telles que la normalisation, la gestion de la sécurité, la gestion, les droits de propriété/d'auteur, etc. et ce du point de vue de l'utilisateur, du fournisseur d'information et du gestionnaire de réseau.

Technical Information Panel

Chairman: Ms Gladys Cotter
Chief
NASA Scientific and Technical
Information Program
NASA Headquarters, (Code JTT)
Washington DC 20546
United States

Deputy Chairman: Dr Maria Cruz Gutierrez
Div. de Materiales y Estructuras
INTA
Crta. Torrejon a Ajalvir, Km.4
28850 Torrejon de Ardoz
Madrid
Spain

MEETING PLANNING COMMITTEE

Meeting Directors:	Mr M. Schryer	(CA)
	Mr M. Brandreth	(CA)
Members:	Mr E.T.H. Lapeysen	(BE)
	Mr B. Hisinger	(DE)
	Mr R. Bernhardt	(GE)
	Mrs D. Patrinou	(GR)
	Mr A.S.T. Tan	(NE)
	Mr P.J. Hoogenberk	(NE)
	Mr Del Rey	(SP)
	Ms A. Özdil	(TU)
	Mr C. Bigger	(UK)
	Mr Walshaw	(UK)
	Mr M.R.C. Wilkinson	(UK)
	Ms G. Cotter	(US)

PANEL EXECUTIVE

Mr G.W. Hart

Mail from Europe:
AGARD—OTAN
Attn: TIP Executive
7, rue Ancelle
92200 Neuilly-sur-Seine
France

Mail from US and Canada:
AGARD—NATO/TIP
Unit 21551
APO AE 09777

Tel: 33(1)47 38 57 95
Telex: 610176 (France)
Telefax: 33(1) 47 38 57 99

Contents

* Paper prepared for the meeting but not presented.

TECHNICAL EVALUATION REPORT

M. VanBuskirk
Canada Institute for Scientific and Technical Information
National Research Council of Canada
Ottawa, Canada, K1A 0S2.

INTRODUCTION

In welcoming the attendees, K. Peebles, the Canadian National Delegate to AGARD, announced that this was the last Specialists' Meeting to be organised by AGARD-TIP under its current mandate. A recent review had recommended that TIP not continue as an AGARD panel. This should in no way be interpreted as indicating TIP's unimportance; on the contrary, its work on the management of scientific information is clearly of the greatest value. Rather, the recommendation reflected the fact that TIP's scope was far broader than aerospace information; hence it fit uncomfortably with the other very technical, focussed AGARD panels, especially in an era of tight budgets. Another location within NATO might be more appropriate to TIP's broad view of the management of scientific and technical information.

M. Scott, the National Librarian of Canada, referred to information as one of the mainstays of research; technologies which support the rapid exchange of information around the world accelerate research and economic development worldwide. High-speed telecommunication systems, originally developed for the exchange of computer resources, are now used principally for information exchange, and dramatic increases in bandwidth as well as system applications are enabling exponential growth in the quantity and variety of information being exchanged and its speed of transfer. Such growth does not occur without problems, and international problems demand international solutions.

This meeting provided an international forum to air some of these problems and suggest solutions. Over 130 people from 15 NATO countries attended, representing academia, associations, industry, research institutes, military organizations and other government agencies. Twelve invited papers were presented from seven countries. A thirteenth paper, though submitted, was not presented as its Russian author was unable to attend. What follows is an overview of the main points from the papers and the discussion, leading to a summary of the significant outstanding issues and some recommendations for how AGARD or NATO might address them.

SESSION I. OVERVIEW

Before we can address the challenges of high speed networks and communication protocols, we must first understand them. The first two papers offered an introduction to these topics.

Paper 1. Global Information Networks: an Introduction to the Internet and its Services

G. Cleveland, a network specialist at the National Library of Canada, described the Internet's evolution from an experiment in computer connectivity in the late 1960's into an international network of networks - an exciting and still rapidly growing mechanism for information exchange and professional communication worldwide with the potential to transform society.

Internet is not an independent homogeneous network but a conglomeration of thousands of separately administered networks, in a hierarchy of local area networks, regional networks and national backbones, all held together by a common suite of protocols. Full Internet connectivity links individuals all over the industrialized world, including Russia, India, Brazil, Argentina and South Africa. Some level of connectivity is available nearly everywhere except in north and central Africa.

Three types of services or applications are supported. Electronic mail includes not only personal communication but also interlibrary loan messaging, electronic publishing and electronic conferences (also variously called discussion groups or bulletin boards). Remote login permits access to a wide range of information resources, such as library catalogues, citation and other types of databases, Campus-Wide Information Systems and Freenets. Finally, file transfer offers many applications for libraries, such as electronic document delivery, the exchange of bibliographic references, and the ability to browse through remote files and copy any of interest.

The phenomenal growth of Internet resources has led to a resource discovery problem - there is too much information and it is too hard to find. This problem has prompted the development of new tools for navigating through Internet, such as Archie for locating anonymous FTP sites, WAIS for handling natural language queries, Gopher for menu-driven access to all electronic resources, and World Wide Web for hypertext access.

Questions from the floor dealt with a wide range of issues. Regarding computer viruses, the level of risk from Internet was considered low. Viruses originate from software, not the network, and reputable FTP sites ensure their software is virus-free. Furthermore, organizations can develop local procedures to defend against viruses and other security problems. On the complex issue of commercialization, it was pointed out that each country has its own policies of access and acceptable use. Even where acceptable-use policies have been strictly against commercialization in the past, these are gradually being relaxed.

Technical Evaluation Report on Technical Information Panel Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

Paper 2. Global Information Networks - How they work

A. Bodini of the European Space Agency distinguished between services provided by the networks and services provided by the hosts. He concentrated on the network services, as host services were the applications addressed during the remainder of the meeting.

He first described the OSI (Open System Interconnect) Reference Model, developed by ISO to resolve the problem of moving information between computers. It breaks the problem up into smaller parts. The model defines seven layers of protocols to define and regulate communication. Each lower layer provides a service to the layer above. The three lowest levels of the protocol stack define the physical aspects, error correction and routing of messages through the network. They ensure that the applications of the upper layers can be developed as end-to-end functions independent of the underlying network. In the Internet environment, the Internet Protocol (IP) defines the same network requirements. The fourth or Transport Layer of OSI is equivalent to the Transmission Control Protocol (TCP) of the Internet. The top or application layer defines the specific services supported by the network, such as file transfer and electronic mail.

Network services comprise connectivity, value-added services and network management. These services are provided differently by different networks, the suite of protocols and standards ensuring that each network can communicate with all others. Protocols such as X.25 and X.28 in the OSI environment or Telnet in the Internet environment ensure seamless connectivity, while OSI's X.400 and X.500 and Internet's FTP and SMTP define specific file transfer applications.

In questioning, it was stressed that convergence of networks and protocols into a single common "operation" remains a goal rather than a reality. At present, each service is single-protocol, but multi-protocol services and gateways offer a partial solution. Much inter-network telecommunication in Europe is done through OSI X.25 circuits; this works as long as TCP/IP is "spoken" at each end of the circuit.

SESSION II. MANAGEMENT

High-speed networks present immense management challenges, at the national and network level as well as at the library and user level.

Paper 3. CANARIE - The Canadian Network for the Advancement of Research, Industry and Education

W.J. Padden from Industry and Science, Canada presented an overview of CANARIE as an example of network management at the national level.

In 1988, Canadian regional networks supporting TCP/IP were established, followed by a research project called CA*net to link them. CA*net has a 56 kbps backbone linking the ten regional networks and three links to the US NSFnet. The target users were industrial and government R&D laboratories, universities and colleges, and libraries and schools. With the evolution of CA*net into a viable commercial operation, responsibility was transferred to Industry and Science Canada and the perspective changed. CANARIE's objectives are to stimulate research throughout the country; to accelerate development of the next generation of hardware, software and services; to accommodate the needs of existing CA*net users while providing a growth path; to expand training opportunities for engineers and other professionals; and to stimulate the development of databases.

To resolve conflicting interests, working groups were established, to deal with business planning, network architecture, governance, and marketing, and later to oversee the upgrade of the regional networks to parity with the national backbone. CANARIE is to be a joint venture, with public as well as private funding, and managed by a corporation at arm's length from government.

The initial phase will see a backbone upgrade to T1 speed or higher, stimulated R&D on lightwave technology and OSI standards, and the development of distributed services and laboratory facilities connected to fibre. The second phase will see an upgrade to T3 speed or higher and an expansion of the core test network, with continued research and development. The third and final phase will showcase the application of CANARIE to the economy at large, by which time it will hopefully be commercially viable.

Questioning pointed out that CANARIE takes a very nationalistic view, that no mention was made of international links except to the USA, and that forcing Europe-Canada traffic to pass through the USA puts a strain on US carriers. However, beyond phase I, nothing is firm; originally government was to gradually bow out of the process, allowing industry to provide the direction. Although the concern was raised that, if CANARIE's direction is commercial, then no international link will ever happen, it was also pointed out that the Canadian and European space agencies are funding the development of switched connectivity which will lead to improved connectivity with Europe. Following a suggestion that CANARIE might be too tightly bound to OSI, which is supported by government but not by industry (where it is seen to reduce rather than enhance competitiveness), it was indicated that CANARIE is softening its position on OSI.

Paper 4. Standards for data and document interchange.

G. Stephenson, from Knowledge Exchange Technologies, Ltd., Luxembourg, described the context in which information standards are evolving. Current technology allows us to move bit streams around, but it is essential that we be able to read the document at the other end, not only today but also in the future. Users use the network for a purpose, and do not want to have to understand how it works and what its problems are. Data transfer and processing must be seamless, the end product must be appealing and usable, and the data must be preserved and readable over time. Standards will contribute to an environment where the information is independent of the processing. These standards are of two types - content and structure.

Content standards define the information at its lowest level. For text, ISO 646 (ASCII) has been used for coding characters, but for 8-bit coding there is no such standard; ISO 10646 for 32-bit coding should allow for the coding of all character sets and all languages. For vector images, a number of competing proprietary formats are impeding standardization, while for bit-mapped images the dominant standard is TIFF, a de facto standard which presents a number of problems. There are also standards for compression of line-art, grey-scale and moving images and of audio to facilitate transmission over low-bandwidth.

Structure or document standards define the organization of the information. For text, two different standards developers have worked from opposite directions. Open Group Architecture (ISO 8613) set out to be a model for all documents so as to be independent of application; it is very complex and rarely implemented. Standard Generalized Mark-up Language (ISO 8879) created a syntax for adding computer-readable information to a document, generally for publication purposes. At the presentation level, the de facto standard Postscript is frequently adopted for lack of a stable international standard. Standards for multimedia and hypermedia are in a similar state of evolution.

There is a need to promote both the existence of these standards and their value in document exchange. Questions related to the distinction between industrial *de facto* standards and international *de jure* standards, and especially, given the long time required to develop the latter, why we should wait for them. It was pointed out that, for international standards, there is a predictable schedule and mechanism for updating, that conformance testing is possible, and that earlier versions are more likely to be compatible with later ones. De facto standards provide short-term solutions at the expense of long-term stability.

Paper 5. The Virtual Library: Coming of Age

J. Hunter of the NASA Scientific and Technical Information Program described how the "virtual library" provides a practical way to start using the elements we have been discussing so far.

A video, produced by Apple Corp. in 1987, presented a view of desktop information retrieval in the 21st century. But much of the technology exists now - high speed networks are in place, distributed computer architecture is affordable, and navigation tools, interfaces and multimedia access are being developed daily. Today we can offer the illusion of one-stop shopping - to locate formal documents, to locate papers and colleagues and contact them using electronic mail, and to locate information documents using network navigational tools.

In the 1980's, the stress was on quantity of information; in the 1990's it is on quality. The information can be owned and controlled by the owner, but readily accessed by anyone, resulting in a reduced need for libraries and researchers to acquire everything 'just in case'. The information can be acquired at the desktop, its physical nature and geographic location being irrelevant, and it can be easily manipulated by the user. However, the virtual library presents many challenges. Such change does not come without cultural cost, as people resist change and fear for their jobs. Nothing is possible without standards, which are under development but generally not finalized. The networking infrastructure is not adequate to cope with multimedia. Intellectual property rights must be addressed. And electronic archiving - ensuring that we can still read data 30 years from now - is a major problem.

The librarian of the virtual library, in an environment where budgets are shrinking and collections are becoming more and more specialized, will be a knowledge manager, organizing information, managing information, and providing access. Is there a need for this supersophistication? Yes, there is a demand now; yes, it is limited, but the demand is coming from those with the money and will snowball. We must know where we are going before we can plan how to get there, and the virtual library is where we are going.

Paper 6. Problems of international exchange of scientific and technical information for a Russian aerospace research institute

In the former USSR, the emphasis was not on the exchange of information but on its classification, not on improved access but on stricter control. This culture, combined with inadequate network bandwidth and poor international connectivity, creates an immense challenge for Russian information professionals. Solutions being attempted by the Central Aerohydrodynamic Institute (TsAGI) are described in the paper, which unfortunately was not formally presented.

SESSION III. APPLICATIONS AND CASE STUDIES

Despite the challenges described above, much is being accomplished right now. The following papers illustrate specific applications of high-speed networks in the information industry.

Paper 7. Electronic Document Delivery - towards the Virtual Library

CISTI, the Canada Institute for Scientific and Technical Information, delivers documents from a large library collection. Despite automated processes for document ordering and order processing, document copying and delivery are still largely manual activities in an environment of escalating demand and diminishing budget. M. Brandreth, head of CISTI's Technology Group, outlined innovative attempts to improve this situation.

CISTI experimented with the Ariel workstation, produced by the US Research Libraries Group, which comprises proprietary software running on specific off-the-shelf equipment. Ariel technology was found to be affordable, reliable, and of high quality. Nevertheless, the slow scanning speed and limited user interface, together with the proprietary nature of the software which precluded customization, rendered the Ariel solution unacceptable for CISTI.

A second trial, with Faxon Research Services of Boston, tested another configuration using off-the-shelf hardware and custom software. The scanner, with a faster scanning time and a flat scanning bed, met CISTI's volume requirements, but the complex communication chain and the limited user interface caused CISTI to continue its search.

A new Electronic Document Delivery System (EDDS) will use off-the-shelf hardware and software for imaging, document management and workflow. It will use client-server technology, and offer the capability of storing scanned images. A client information file will store, as well as address and financial data, the client's preferred method for receiving documents, and the system will transmit documents in X.400, TIFF, or SGML formats, as well as by fax or regular mail. Incoming orders will be call-numbered automatically; the number of pages scanned will be recorded for invoicing and copyright purposes; and tracking and statistics functions will permit close monitoring of the entire process.

Since much of the delivery of documents will be over the high-speed networks, the networks themselves introduce some challenges. Historically, CISTI has absorbed telecommunications charges for its services; however, incremental use of the Internet appears virtually free to users, who may demand differential pricing. Clients demand a highly reliable service, but with the hierarchical management of the high-speed networks, no one organization is responsible for quality and error recovery. The form of submitted orders is not standardised, unless clients are ordering from CISTI's own databases or using EDIFACT coding; non-standard orders must be entered into the system manually, presenting a workload problem. Finally, the copyright issue is serious, and not yet resolved - as is frequently the case, technology is far ahead of the legal and administrative machinery.

Questions related predominantly to CISTI's plans to store scanned documents, whether in database or CD-ROM form or even through initial acquisition of documents in digital rather than print form. It was stressed that the outstanding issue is copyright, since current legislation does not permit storing scanned documents. Copyright legislation differs from one country to another; for instance, in the UK, the law prohibits electronic storage, but not in France. It is not authors who want royalty revenue, but publishers; much of this electronic delivery problem will disappear when print publishing and digital scanning are removed from the loop, leaving authors publishing in machine-readable form documents which can be retrieved using file transfer.

Paper 8. La Fourniture de Documents Electroniques en Europe : L'expérience de l'INIST

INIST, l'Institut de l'Information Scientifique et Technique in France, was established as a multidisciplinary centre for the creation of databases for the dissemination of scientific and technical information. It handles approximately 600,000 document orders per year. C. Lupovici, director of production, described how document delivery has been automated at INIST.

Two preliminary European projects provided useful base information. The TRANSDOC project showed that, although satellite transmission of digitized text was difficult and costly, digital scanning and storage on optical disk was feasible. The ADONIS project proved that a digitized graphic image could be integrated into scanned text and the whole incorporated into a database.

From these two projects came the operational project SAN (Système d'Archivage Numérique), in which 1800 of the most heavily demanded titles (out of the total collection of 27,000 titles) were systematically scanned and the images stored on double-sided optical disks in a jukebox. The project was intended to answer questions related to availability of the document for delivery, quality of reproduction, speed of transmission, extent of automation and reduction of storage space.

Orders can be received via information retrieval systems such as Questel, ESA or Dialog, from ILL systems such as PEB in France or OCLC in the USA, by direct capture from the INIST online catalogue, or in non-electronic forms such as fax, paper or telephone. Orders for documents stored on optical disk are handled by the SAN, which processes them every two hours. This period balances the need to sort orders by physical address on the disks and the need for rapid service by clients. This electronic management system offers enormous advantages in regulating workflow, generating statistics for collection development and for selection of titles for scanning, and automatic generation of copyright reports.

A number of problems have emerged. The cost of the project is high. The original list of high-demand titles was derived from existing data, but in fact 60% of the list has changed. Graphics in the scanned documents take up a significant 25% of the disk space. And it is worth noting that rarely is an article requested more than once, and many are never requested at all. This result was something of a surprise, and presents as much of a problem to journal publishers as it does to INIST. In fact, publishers have tended to prefer copyright royalties to high subscription costs; this experience might indicate that the reverse is a wiser choice. In the long term, the best solution is for publishers to provide electronic text in the first place. INIST has the capability of handling such technology. What is impeding this is the capability of publishers to modernize their processes and, even more important, their willingness to cooperate in solving the copyright dilemma in the environment of electronic dissemination.

A question on the proportion of Japanese titles being digitized led to an admission that literature from all non-Latin alphabets is underutilized, food for thought in a meeting on global information networks.

Since INIST informed publishers about the electronic storage of documents, several have requested special agreements. INIST has an agreement with the French copyright centre covering all document supply, domestic and foreign. Again, it was pointed out that copyright is a problem for publishers, not for authors. And it was emphasized that the copyright problem must not stand in the way of attempts to solve the technical problems; otherwise neither the information centres nor the publishers will survive the technical revolution.

Paper 9. Electronic Journals

The production of journals in electronic form would get around many of the delivery problems illustrated in the previous two papers. C. McKnight of the HUSAT (Human Sciences and Advanced Technology) Research Institute in the UK, examined the status of electronic publishing.

With all the problems of paper journals (misshelving, binding, in use by someone else, hard to search through, take up too much space), and the advantages of electronic journals (always available, easy to search, take up almost no space), it is useful to ask why electronic journals have not proliferated, because it is these issues that will have to be addressed. A number of experiments with electronic publishing in the UK and the USA have provided useful insights.

First, electronic journals are not so portable or user-friendly as print products. Then, the graphics offer neither the quality of print nor adequate portability across platforms. Internet is still not universal (in the UK it is chiefly in the academic domain), so exclusively electronic distribution would not reach all potential subscribers. Articles appearing in electronic form are often not considered acceptable for performance review and tenure; for this, articles must be peer-reviewed. The entire costing model for electronic journals has not been resolved, nor has the question of copyright control. Electronic publishing alone is not enough; we must be able to do at least as much with an electronic journal as with a paper journal.

Questioned on the relative importance of these factors, it was suggested that it is not the technology that is holding up acceptance of electronic journals; the publishing industry needs to look at itself critically to find alternative models, as libraries are, or go out of business. Also publishers must cooperate with each other if we are to avoid many incompatible interfaces. Two other suggested problems with electronic journals were the difficulty of citing references (e.g. what is a page?) and the question of permanence. An electronic record is as permanent as a paper record, but the method of reading it is not.

Paper 10. Bulletin Boards, Electronic Mail, Conferenceing. Current use by Scientists and Engineers; Effects on Libraries and Information Centres - Do they have a Role?

A bewildering array of electronic mail, conferencing and file transfer resources is available, as well as a range of applications (such as Gopher, WAIS and World Wide Web) for locating and using them. V. Castelo, Communications Manager of the Consejo Superior de Investigaciones Cientificas (CSIC) in Madrid, examined some of these in the context not only of information-seeking by scientists and engineers, but also of the role of the librarian in the process.

Scientists appreciate the convenience (services are all available at the desktop), the opportunity to perform the search without an intermediary, the speed of response, and the ease of informal communication with peers. For librarians, these changes can be seen to pose a threat, since researchers may no longer require their services. Nevertheless, librarians do have an important role, and must seize it. In the new environment, librarians should become familiar with the technology and be prepared to train researchers in its best use (at present, this role is being taken by Computer Centres, a situation which ought to concern librarians). They should take a role in updating and organizing the information available. They should also become involved in the standardization and testing of products to ensure quality and ease of use. While encouraging the researchers to perform their own simple searches, they can provide support for complex searches, and maintain files of contacts and resources. Finally, they can perform an important role in filtering, assessing and controlling quality, and adding value. In short, librarians should participate in the design, development and maintenance of applications and resources on the networks, cooperate with computer centres, and integrate the new technology into the existing information flow.

During the discussion period, it was stressed that librarians must make the leap and learn the new technologies or they will be by-passed. Searchers and librarians must continue to work together. It was also suggested that informal information-gathering be distinguished from formal. In the former, such as a researcher's use of electronic mail instead of the telephone for collegial communication, there should be no need for the "interference" of a librarian, especially as the engineers will probably learn the technology first and then train the librarians. Document delivery, though, is a very promising application in which librarians will clearly have a role.

Paper 11. Training and Operations Support System (TOPS)

A different application of high speed networks is their use in real-time training and problem-solving. I. Gordon of CAE Electronics in Montreal described a hypothetical system for on-demand access to distributed training, which could include online documentation, schematics, video, audio, 3-D intelligent models, operating procedures, and simulations. High-quality screen images can exploit simulation techniques, colour coding, flow indicators, zooming, panning, and hypertext links to maximize the transfer of information in the least time. He stressed that all of the parts exist now; it is only the whole which does not yet exist.

TOPS has three major components - a front-end with an intelligent, friendly and appropriate interface, a bank of media-rich data sources, and servers with powerful database management systems for storing, retrieving and interpreting the data. It relies on three enabling technologies - high speed networks for timely data transmission (100 Mbps to the desktop), object-oriented database management systems that can manage complex information and its dependencies across distributed systems, and high performance hardware and software.

The scenario described in the paper is in the field of aircraft maintenance, but TOPS is appropriate for any field where the acquisition and maintenance of information and skills is complex and costly, and where staying up-to-date is critical.

Questioned on a detail of the scenario in which the aircraft maintenance engineer appeared to make a change to the procedures, the author clarified that in fact the engineer would merely tag the procedure; the system could then flag such problem procedures for future handling using the appropriate change mechanisms. Is there a demand for this application? Yes, at least for some parts of it. As with the virtual library, demand for this kind of technology will escalate when its capabilities are seen and tested.

Asked if there were any studies on the effectiveness of this kind of training relative to others, the author pointed out that we are moving toward proficiency-based training, where the manner of the training is less important than the ultimate proficiency. The costing and billing of this kind of training needs to be examined, since it is costly to develop and it is important that developers receive a return on their investment.

SESSION IV. THE FUTURE

The future is a relative term - for some it is a long way off, and for others it is tomorrow. Nevertheless, the last two papers attempted to project what is now seen to be possible in the very near future in terms of delivering scientific and technical information.

Paper 12. Wide Area Information Servers: an Executive Information System for Unstructured Files

WAIS, or Wide Area Information Server, already exists as a navigation tool on the Internet. B. Kahle, president of WAIS Inc., described the genesis of WAIS and how it will continue to evolve to simplify access to Internet information.

For the typical engineer, information can be organized in a hierarchy from personal information through organization information to published information. Although personal information ranks highest in terms of personal importance, it ranks lowest in terms of quantity and ease of access, while published information, least in demand, is easiest to access.

The 1970's were the decade of the mainframe, and the 1980's of the personal computer and LAN. The 1990's are the decade of the Wide Area Network (WAN), with three important implications - the client server model, international standards and logging for accounting.

WAIS is a navigational tool optimized for the WAN, offering access to personal, organizational and wide area information in an interactive mode. It offers mixed natural language and Boolean searching, accounting for billing purposes, ability to restrict access and ensure security, provision of data in its original form, all running on the popular UNIX platform. The WAIS protocol supports NISO Z39.50, in a client-server environment. To be working effectively, the WAIS client should be busy 24 hours per day, scouring the world for information, gossiping with other users, pondering retrieval. The client software exists as a separate and individual copy on each personal computer, and communicates using TCP/IP on the LAN.

WAIS is particularly useful in accessing unstructured and semi-structured data such as data archives. However, personal data tends to be very difficult to access using WAIS. Questioned on data control, the author stated that publishers can control data by keeping it on their own servers, and that they can restrict and/or charge for access, and ensure observance of copyright regulations. Asked about the role of the librarian, he used the analogy of accountants, who were empowered by computer analysis and brought out of the backroom and into the boardroom; he sees the same possibility for librarians. To make an existing database WAIS-searchable, it is simply necessary to connect the computer to a WAIS server, which will automatically keep the index up-to-date, leaving the data itself in the original computer files. WAIS is for access, not storage.

Paper 13. Information Retrieval Services on Wide Area Networks: A Vision of the Future

Internet is only one of many networks used to access and transfer data. J. LeMézec of France Telecom North America in New York City looked at other networks with a view to pointing out areas where development work is needed.

Despite the emphasis placed by the US government on the Internet and the High Performance Computing and Communication (HPCC) Program, the telecommunication network providers are developing X.25 data networks and committing to CCITT and OSI standards rather than to TCP/IP. Interfaces permit communication between the two types of networks but there are difficulties. In France, the VideoTex service TELETEL (also known incorrectly as Minitel, the name of the terminal) has proved a success, largely because of the attention paid to human factors, a lesson to other network providers. The network architecture could be used as a model for future information networks. The entire system is in constant evolution, moving toward commercialization of the terminals, high speed access and ISDN applications. Speed upgrades make graphical user interfaces feasible and Audiotex is also available to provide vocal access to remote databases.

France has upgraded its telephone network to ISDN, which enables transmission of graphics and photographs. Existing applications are in sales transactions, tourism, entertainment, distance learning, medicare, meteorology, and scientific and technical information retrieval. Multimedia, the inclusion of other media along with text and graphics together with appropriate synchronization and manipulation, is not yet available, but is feasible in the short term although the standardization issues must be resolved.

In order for an international network of networks to evolve and operate efficiently, it will be necessary to recognize that different kinds of networks co-exist in different countries, and that they are operated differently in every case. Chief among the challenges are the very high infrastructure costs, ensuring the interoperability of heterogeneous networks, privacy and security, intellectual property rights, and control of traffic. The development of navigation tools and user-friendly interfaces is essential for success.

Questioned on the reasons why Internet should have done so extraordinarily well compared to ISDN, the author replied that, in fact, ISDN is being used successfully in France, and that the two do communicate, so their relative success is immaterial.

Invited presentation

N. Goldmann, of Array Development Inc. of Ottawa, volunteered an impromptu presentation on repositioning for the new knowledge economy. Global knowledge as the basis for competitiveness puts tremendous pressure on the information infrastructure. It is tempting to think of the Internet simply as a knowledge pipeline - we just connect to it and turn it on. But it is not as simple as that. High-speed telecommunications represents not just an incremental change in technology but a paradigm shift changing the role of all players. And as with all paradigm shifts, there can be no correlation between past and future performance and no safe niche. We will see a reconcentration of power and new principal players; for example software publishers are already carving out a role for themselves. Old institutions will simply move to the new paradigm or die.

In particular, conventional library-oriented methods of information dissemination are inadequate, and print publications cannot support the transfer of critical technologies from academia to industry. Emerging electronic technology, especially the Internet, is offering scientists alternative ways of accessing information. Already, many researchers refuse to work without Internet access. It is a myth that users need documentation and training. Most users have taught themselves, and the Internet has expanded and prospered despite the lack of controls and training. Internet brings the invisible college to the desktop, presents a unifying global force, and permits a new way of scientific publishing for rapidly evolving technological fields that makes ongoing discussions international and immediate.

To survive at all, the knowledge worker will need to adapt to this new environment. As Internet emerges as the primary knowledge conduit, the knowledge worker will first have to become connected to and familiar with the Internet and its resources, then have to find ways of obtaining, repackaging and disseminating information in unique and specialized ways; otherwise the end-user will get it for himself.

In the new knowledge economy, knowledge will be the key factor in competitiveness. Organizations will need to find a framework to reposition themselves to function in it. They need to define what they mean by success; for many, this will be survival. Then they will need to build on core competence of their personnel, redefining their strategic mandate and mission.

Open forum discussion

Topics for the open forum discussion were solicited from attendees, and asked in sequence, with all attendees encouraged to respond. In theory, this should have led to a lively discussion of issues; in practice, this was simply a question and answer session. Some of the questions were very basic; some related directly to a specific paper and could have been better handled during the discussion period following that paper; and some reflected complex issues that require time and attention to answer, even assuming there is an answer at all.

Something more structured should be attempted in future, perhaps with like questions grouped into major themes, perhaps with attendees and speakers notified in advance, providing the opportunity to provide reasoned answers and thoughtful discussion.

The questions are paraphrased where necessary, and the discussion summarized.

Why did the Videotex service Minitel succeed in France when similar projects did not succeed in other countries?

There are many contributing factors. French regulations are not so restrictive as those in the USA, for one thing. The whole market situation is different. The terminal was very simple to use. Another reason, of course, was that Minitel was free, and that it initially replaced telephone directory inquiries. Other features were tried out and added gradually.

What proactive steps could AGARD-TIP take to solve intellectual property and copyright barriers in order to maximize the role of international value-added networks?

Copyright is a big issue that keeps coming up. The term, however, applies to the physical presentation of the work ['rights to the *copy*'] and the concept is now stretched. With electronic media and with joint authorship from many countries under different copyright laws, the definition will snap. The concept of copyright will remain important for some sectors (for example, novelists), but for our sector, it will simply disappear. Authors want dissemination, not control and not royalty payments.

What skills do information specialists lack to undertake the consultant role described by Brewster Kahle? Is there a role for AGARD-TIP here?

It is not clear that librarians are lacking skills. Librarians have always investigated new technologies as they come along and adapted those that are seen to be useful.

Are there 3 or 4 concrete examples of how services on the networks are charged for and paid for? What criteria are used and what are the different charging scales?

Again, it is important to distinguish the network from the services on the network. Who pays for the *network* varies from country to country. In Italy, for example, there are two Internets, one created by universities and one by the national institute for nuclear physics and each is financed separately and differently. One of them operates as the national Internet organization and takes responsibility for the administrative work. Their attitude is - if you want to use it, go ahead and use it, but if it doesn't work, we do not take responsibility. You are not charged for use, but the network is not free; someone else is paying. In France and the UK, there are central operators that take this responsibility, in each case funded by government.

There is also the question of who pays for the *services*. Some services are really free, such as OPACs, and anonymous FTP; these were developed principally for internal users, and it is easier to make them available free to all than to develop a billing structure. Other services, such as ESA, must be paid for by the user, who must have a subscription or a contract. It is interesting that all European countries have good connectivity to the USA but not to each other. Perhaps AGARD-TIP could work on this, by encouraging a single national organization in countries where it does not exist and by encouraging bilateral or broad connectivity.

A conflicting view was that AGARD-TIP should not encourage a single national controlling organization; Internet activity in the UK has increased dramatically since the Janet cartel was broken. There should be as many Internet groups in a country as possible.

Is there a single Internet standards body? Will Internet adopt the ISO approach of having formal standards? Who will be responsible for Internet standards?

There are de facto standards developing all the time; TCP/IP itself is a de facto standard. In the Internet Society, there is a process called RFC (Request for Comments). A proposal for a standard can be put up on the Internet as an RFC, and anyone can comment on it. After a period of time, it can be tried out and implemented, and it may become an Internet standard. This occurred with SMTP, with the conversion standard between e-mail and X.400, and with domain names and naming conventions. However, some RFCs are just for information. There is also the ISODE consortium, which develops OSI applications, such as X.400 and X.500 running on TCP/IP and a variety of ISO standards.

In research, people experiment and develop new technologies without waiting for standards to be developed and implemented. This approach is one of the reasons why the Internet is so successful. Standardization comes later, after the dust settles.

Do journal publishers have any vested interest either in promoting the academic acceptance of electronic journals or in resolving the citation problems ? What is the role of book dealers and subscription agents in electronic publishing?

Publishers have a vested interest in staying in business. They all want to get into electronic publishing, but they also want to make money and none of them wants to make the first mistake. That is why Oxford University Press entered the market with the Oxford English Dictionary, a product with which they simply could not lose. It is significant that many publishers have got rid of their printing houses, and are contracting out their printing. Apart from maybe bundling favourite electronic journals on CD-ROM, it is doubtful that subscription agents will have a role in electronic publishing.

Yes, publishers are ready to take the second step. TULIP posed no technical problems at all; the problems were with the publishers and users and were all non-technical - such as pricing structure, the transitory nature of the document. Publishers will simply have to change. The place of subscription agencies in electronic publishing is not obvious, but they will be trying to find a place for themselves; perhaps it will be in table of contents-type services or document delivery. FAXON, EBSCO and SWETS are already trying this. But clearly their traditional role will not be sufficient to survive.

What does 'high speed' mean? What bandwidths and what equipment are needed for specific services? For example, net news requires 19.2 kbps and a *workstation.*

Low speed is defined as anything under 64 kbps. High speed is anything above that. To give a specific example, in the handling of remote sensing data, 64 kbps is adequate for the acquisition of low-resolution images, but the browsing of these images requires 256 kbps. However, the delivery of a 60 Mb image requires up to 2 Mbps; this is so costly that physical transport is preferred at this time.

Isn't the MOSAIC interface an integrated way of accessing Internet resources such as WAIS, Gopher, WWW)?

Yes.

What is the level of traffic on Internet and does it impact on operations?

The speed of the Internet depends on the country and the speed of the connectivity between the user and the host. Between CA*net and the US and between the USA and Europe the links are 1.5 Mbps. One of the major traffic blockers is multicasting, which constitutes 30% of the traffic. Speed depends on the time of day and the load on the host systems, especially on those which are free. In any case, CPU cycles on the host are a more expensive component of the search than the speed of the network.

How can we promote the virtual library to traditional users who find the environment threatening, without duplicating our resources electronically and in print?

We are not talking about drastic changes in service; we are talking about integrating new technologies into traditional services. Gateways and scanned documents merely expand service and reduce cost. We should be keeping the traditional and adding the new if it meets requirements.

Many North American industries are working towards ISO 9000 as a standard for compression. Will ISO 9000 have a role in the quality of standards development and compliance on the Internet?

The quality of an individual standard is not in question. However, ISO 9000 might help in the ISO certification and in assuring the quality of an implementation.

In Canada, there are at least two ILL products which purport to comply with the ILL protocol but in fact do not comply fully. What role does CISTI play in ensuring that software complies with the ILL protocol?

In Canada, it is the National Library of Canada which takes the leading role in the development of standards for library applications and in testing for standards compliance. CISTI is primarily a production organisation, making the best use of the standards that are available.

How can the quality of data on the Internet be assured?

With an edited electronic journal, there is the system of peer review. Otherwise, there is an enormous amount of information whose quality is completely uncontrolled. In one organization where the role of indexers is under review since indexing is becoming more and more automatic, the use of indexers (who are highly trained in science) for filtering and validating data on the Internet is being looked at.

COMMENTS

Thirty-four of the attendees completed an evaluation form. Almost all of them indicated that the meeting was worthwhile and successful. Topics were of interest, thoroughly discussed at about the right level, and most of the speakers were judged effective in their presentation, about half judged to be of high quality and the other half of average quality. The physical arrangements of the meeting were all rated highly.

One respondent asked for demonstrations of prototypes and/or products to illustrate the concepts, another for more case studies to assist in day-to-day problem solving, another for a thorough state-of-the-art report on the design and implementation of network-based information services.

The international approach was cited by one respondent as being useful, creating a feeling for the entire international community and where it is going. Another mentioned a pressing and continuing need for educational and information-missionary activities in the area of high-speed information networks.

As mentioned above, the purpose and handling of the Open Forum Discussion Period needs to be re-examined.

A number of suggestions for follow-up meetings were offered:

- a review of core competencies for information specialists in the 21st century;
- actual use of networks by information centres and their impact on the centre's organization, products, parent organization, etc;
- examples of value-added information services by information centre staff - types, trends, markets, clients, value;
- MIS (management information services).

These all suggest a need for practical assistance at the level of the information centre. While there is no denying the need for this type of help, it is not necessarily true that AGARD-TIP or its successor is the appropriate vehicle for it. There are many national and international conferences dedicated to library and information centre applications. Perhaps AGARD-TIP should concentrate on areas where international understanding and cooperation can achieve something unique.

MAJOR ISSUES

As with other issues in our society, the technology of high-speed information transmission is moving ahead, forging new directions and challenging our attitudes, with the societal (cultural, legal and management) aspects lagging behind. From the papers presented at this meeting, outstanding issues and challenges coalesced into four broad areas:

1. There is a serious need for an international perspective and an exchange of information at the international level about national high-speed networks at both the technical and application levels. This includes topics such as connectivity and infrastructure, security, services available, management and funding, commercialization and charging for services, roles of government and industry, etc.

2. Standards are an important and continuing issue. International understanding of the standardization process itself, the bodies involved, the various players and where OSI fits in, standards for convergence of networks, for communications between networks, for services and their portability across networks, for archiving and preservation of data, for electronic publishing interfaces, all these and more are critical and ongoing.

3. Copyright is a recurring problem, at an international level since we are talking about international information exchange. The various national regulations on scanning, on store and forward activities and on handling of information that exists only in electronic form need to be fully understood, as do the economics and attitudes of publishers. New mechanisms for analysis and recovery of costs need to be explored, as well as the real issues of crediting innovation.

4. Finally, and perhaps principally, the new environment presents a culture change, even a paradigm shift. All players from author to end-user must come to terms with the environment and its impact on themselves and all the other players. Authors need to be assured of receiving credit for new ideas, and that their work in electronic form will receive the same degree of attention and be properly cited; in academia, the use of peer-reviewed journal publication as a performance indicator for tenure needs to be addressed. Information providers (including traditional print publishers, online vendors, and others) need to learn how to operate successfully, how to earn revenue, recover costs, and carve a market niche. Information professionals need to find new roles in testing technology, navigation, filtering, training and adding value to information. End-users need to solve problems of volume and quality of information.

"As usual, the solution of these problems will require a balance between regulations and liberty, between competition and cooperation. The presence of everyone here is an example of cooperation." [LeMézec]

Because these are all international problems, these are all topics which can be fruitfully addressed by the successor to AGARD-TIP.

RECOMMENDATIONS FOR AGARD-TIP

As the future of AGARD-TIP is being debated and assessed, it is useful to consider exactly what was achieved in this meeting that could not have been achieved in another forum. It was the bringing together of international experts from government, academia and industry, representing service and information providers, network experts and consumers, and from many different countries world, that was unique. Information is an international commodity, and the opportunity to discuss issues, solve problems, exchange ideas, forge and maintain contacts, all at the international level, is of critical and growing importance. Many conferences and meetings can and already do give a local or regional picture; only a meeting such as this can provide the international picture.

Aerospace information is only one of many types of information, and even specialists in one area need information from other areas, so it does make sense that TIP be permitted to keep its horizons broader than aerospace. If TIP remains located within NATO dealing with the management of most of scientific and technical information, it would be well if some attention were also be paid to the needs and concerns of the developing world, and to the handling of information in non-Latin languages.

AN INTRODUCTION TO THE INTERNET AND ITS SERVICES

by

Gary Cleveland
National Library of Canada
395 Wellington St
Ottawa, Ontario K1A 0N4
Canada
gary.cleveland@nlc-bnc.ca

Abstract

The emergence of research networks——computer networks that serve the academic and educational communities——offer libraries throughout the world the opportunity to enhance library resource sharing and information access. The best known example of this structure are the networks collectively known as the Internet. The Internet is actually a conglomeration of thousands of networks found in North America, Europe, and every other continent. It is a "metanetwork" that: 1) physically interconnects TCP/IP-based networks, many of which have hundreds of component networks; and 2) provides gateways to other types of networks, such as BITNET, an electronic mail network that supports computer communications through 32 countries. This paper provides a non-technical, high-level overview of the Internet and its services. Specific topics include a description of the Internet, how it began, and its current size; its overall structure; the types of international networks it interconnects; the network services that are available; and how these network services are used by libraries.

1. Introduction

The Internet——a vast "metanetwork" that connects thousands of networks and over two million computers worldwide——has been causing much activity and excitement throughout the international library community. Just a few short years ago, not many librarians had heard of the Internet, but now many articles, conferences, and workshops are devoted to the Internet each year.

The library and information communities are enthusiastic about the Internet because it offers, for the first time, concrete step toward creating desktop access to information——the so-called "electronic libraries" or "information networks" that librarians have envisioned for many years. The Internet offers access to a wealth of information that includes hundreds of library catalogues, citation databases, electronic conferences, electronic journals and newsletters, satellite images, archives of text, images and software, and a growing number of sophisticated resource discovery tools such as WAIS, Archie, Gopher and World Wide Web. Clearly, the Internet has become a valuable tool for information access and professional communication.

2. Origins of the Internet

The Internet was originally created in the U.S. in the late 1960's as an experiment in computer connectivity. It was a project sponsored by U.S. Defence Department's Advanced Research Projects Agency (ARPA). This early network——called ARPANET——was a wide-area, packet-switching network that initially connected just a handful of computers at a few sites around the U.S.

During the 1970's, this infrastructure slowly grew, splitting into two networks——ARPANET and MILNET. Throughout this period, TCP/IP, the set of telecommunications protocols that has come to define the Internet, was developed. TCP/IP stands for Transmission Control Protocol/Internet Protocol. It is the common "language" of the Internet that allow computers to communicate.

In the mid-1980's, the US National Science Foundation (NSF) created a network based on TCP/IP called NSFnet——a large coast-to-coast network that interconnected many smaller networks and provided researchers with access to resources such as supercomputers and specialized software. With the involvement and funding of NSF, this proto-Internet slowly began to connect more and more networks and computers located at research institutions, universities and colleges, government departments, and research-oriented private industry.

Until recently, the Internet was the exclusive domain of these large research and educational institutions. However, there has been a dramatic shift in the last three years in the types of groups that access the Internet. Currently, smaller organizations, commercial companies, and even individuals are gaining access to the network. In fact, commercial organizations now constitute the largest and fastest growing group on the Internet. This expansion of the Internet to include more diverse users groups is the result of a number of factors:

- the easing of restrictions on commercial traffic on the Internet. At one time, blatant commercial activity was prohibited, but these restrictions are slowly being relaxed.

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

- the emergence of private, commercial TCP/IP networks that sell access to the Internet.

- the development of low-cost methods of accessing the Internet, such as casual, dial-up connections.

In 1993, the Internet is now a vast "network of networks" that physically interconnects thousands of networks worldwide. It has evolved from an experimental network, to a research-oriented network, and now to an open, global network that provides access to thousands of diverse information resources and services.

3. Internet growth

Since about 1985, the Internet has been growing at an extremely rapid rate. In 1981, for example, there were just 213 hosts connected, and by July 1992, there were well over 1,800,000. By the end of 1993, the number of individual networks is expected to grow to over 11,000. Though the number is difficult to enumerate accurately, the total number of individuals accessing the Internet is estimated to be 35 million.

4. Internet structure

The Internet itself is not a separate network in the sense that it is an independently administered, homogeneous entity. Rather, it is a conglomeration of thousands of separately administered networks. It is a "metanetwork" that: 1) physically interconnects TCP/IP-based networks, many of which themselves have hundreds of component networks; and 2) provides gateways to other types of non-TCP/IP networks.

The networks connected to the Internet typically have a three tier structure. The first tier is comprised of local area networks (LANs) that provide connectivity among computers at a given institution. The LANs, in turn, are linked together through regional networks——the second tier in the hierarchy——to support communication among institutions throughout a particular region. An example of a regional network is NYSERnet, a network that provides connectivity throughout the Eastern US. The third tier is the national backbone network that interconnects all regional networks into a national system. For example, NYSERnet is connected to NSFnet, the US national backbone network.

While the Internet connects thousands of different types of networks, scientific and technical information professionals should be aware of three distinct types: 1) national backbone research networks; 2) international networks; and 3) non-TCP/IP networks.

National research networks
Almost every major industrialized country possesses a national research network. These networks are based on TCP/IP or, for multi-protocol networks, TCP/IP is one of many telecommunications protocols supported. Examples of national networks include:

- ***NSFnet*** in the United States
- ***CA*net*** in Canada
- ***AARNET*** in Australia
- ***JANET*** in the United Kingdom
- ***NORDUnet*** in Scandinavia

International networks
There are also a variety of international networks connected to the Internet that span many countries. Major examples include:

- ***DDN***. This is the U.S. Defense Data Network operated by the Department of Defence. It includes MILNET and a number of other classified military networks.

- ***NSI***. The NASA Science Internet combined two earlier networks, the Space Physics Analysis Network the NASA Science Network. The current network, NSI, supports TCP/IP and has nodes on most of the continents in the world.

Non-TCP/IP networks
Another group of networks are those that are not physically connected to the Internet, but rather have gateways to it for the purpose of passing electronic mail. Major examples of this type of network include:

- ***BITNET***. BITNET is a worldwide network that is as old as the Internet that supports primarily electronic mail.

- ***USENET***. USENET is a large global network of Unix machines that supports electronic conferencing and news.

- ***UUCP***. UUCP is another network of Unix machines that began in the U.S., but has since spread throughout the world.

- ***FIDONET***. FIDONET is a simple, low-cost worldwide email network based on telephone connections and modems linking primarily DOS computers.

5. Network services

The best way to organize the uses of the Internet is by the standard network services that support them. There are three standard network services——or tools——that have been traditionally available on the Internet: electronic mail, remote login, and file transfer. These traditional services are discussed below.

Electronic mail
Electronic mail involves sending messages from one computer to another in a *store-and-forward* manner, where the message is relayed from computer to

computer. Once at its destination, the message awaits the addressee until he or she has time to read it. Electronic mail is the most commonly supported service across the diverse set of networks that are connected to the Internet. Where electronic mail crosses a boundary between networks using different protocols, a gateway translates the message into the appropriate format.

Electronic mail supports a number of information-related services which include:

- ***personal communication***. One-to-one email supports personal communication and information exchange among colleagues.

- ***electronic conferences***. A popular application of electronic mail are electronic conferences, also known as discussion lists. Supported by LISTSERV software, electronic conferences have become powerful tools for information exchange, professional communication and development. There are discussion lists available on the Internet on hundreds of topics.

- ***electronic publishing***. LISTSERVs can also be used to support electronic publishing and document access. A growing number of electronic journals and newsletters have appeared on the Internet in recent years.

  One drawback of using electronic mail for electronic publishing and document access is that, at present, Internet mail supports only strict ASCII text. Thus, no complex formatting (e.g., graphics, colour, special fonts, bold, underline, and page formatting) or photographs can be included. However, this limitation is likely to be eliminated with the gradual replacement of SMTP with MIME, a protocol supporting multimedia electronic mail.

- ***ILL messaging***. Electronic mail is also used for sending information regarding the loan of books between institutions.

Remote login
Known as TELNET within the TCP/IP protocol suite, remote login allows a user to connect to a remote computer and interact with it as though his or her local computer were a terminal of that remote machine. In theory, TELNET provides users with the capability of connecting to any system on the Internet. However, security and authentication strategies generally prevent this from occurring. Typically, a user must have an account, a user id, and a password to connect to the remote machine.

There are a variety of information systems available on the Internet that can be accessed via TELNET, both with and without passwords. These include, for example:

- ***online public access catalogues*** (OPACs) of college and university libraries. There are currently hundreds of such systems connected to the Internet. These systems can be openly accessed are generally free of charge. Some public library OPACs have also begun appearing on the Internet.

- ***citation databases***, such as OCLC's EPIC Service, or DIALOG. These systems require an account and a usage fee.

- ***Campus Wide Information Systems***. These are campus-based systems that provide information about a wide-range of topics such as course calenders, the weather, campus events, library OPACs, and are free of charge.

- ***Archie, WAIS and Gopher clients***. While the client portions of these systems typically reside locally, they are also available for use on remote hosts (e.g., a WAIS client is available by TELNETting to *quake.think.com* and logging in as "wais".

File transfer
File transfer, known as FTP (File Transfer Protocol) on the Internet, involves the exchange of computer files between computers. FTP supports the transmission of both binary and ASCII files. Thus it can be used to send, for example, bitmapped images, documents formatted with word processors, executable programmes, spreadsheets, as well as plain text. While there are many general applications of FTP, the three of interest to libraries include:

- ***anonymous FTP*** An application of FTP, called *anonymous FTP*, has become very common on the Internet and represents a valuable method of accessing a distribution information. Under normal circumstances, to transfer files a user would be required to login to a remote machine with a unique username and password. However, many computer administrators have set aside areas on their computers which can be accessed anonymously——that is, without an account on the machine. Such "anonymous" FTP allows anyone on the Internet to login to, browse, and download files. Literally hundreds of anonymous FTP sites, or "servers", exist on the Internet containing files of every description.

- ***electronic document delivery***. FTP is also used for electronic document delivery, sending documents to users in electronic form instead of using fax or regular post. Research Library Group's Ariel Document Imaging Workstation, for example, uses FTP to send journal articles in bitmapped form.

- ***exchange of bibliographic records***. Cataloguing records can be sent via FTP through the Internet as an alternative to magnetic tape through the post.

6. Resource discovery tools

In the past three years, the set of basic services available on the Internet has been expanded to include others designed to help people navigate throughout its thousands of information resources. These tools were created in response to the resource discovery problem, which, in a nutshell, involves the fact that there is so much information on the Internet that it has become very difficult to find information on a specific topic. The situation has changed considerably over the last three years with the development of new tools to help navigate the Internet. The four most popular resource discovery tools are discussed below.

Archie

The Archie system, developed at the University of McGill in Montreal, was created to overcome the problem finding the burgeoning number of files in anonymous FTP sites. Originally, to find a particular file, a user needed to know, first, that it existed, then the address of the host and the directory where it resided, and finally its filename. Archie sites store a regularly updated list of files on FTP servers, which users can search using keywords. Archie responds with a list of FTP sites that contain relevant files.

WAIS

Wide Area Information Server (WAIS) is client-server software developed by Thinking Machines Corporation that assists users in discovering electronic documents without regard to where they actually might reside on the network. In operation, the user enters a natural language query, and then selects a set of WAIS-formatted databases to be searched. WAIS checks each of the databases for relevant files, and produces a list of "hits" that can be viewed or send to the user's workstation.

Internet Gopher

Gopher is a simple, but powerful system that provides easy, menu-driven access to many network resources. A client-server system, gophers provide access to diverse Internet resources including:

- the gophers of other institutions
- anonymous FTP sites
- electronic directories (e.g., WHOIS)
- interactive systems (e.g., library OPACs)
- Archie systems
- WAIS systems
- LISTSERV archives.

Using gopher, users are freed from having to know locations of resources, machine addresses, or even the type of procedure being used. There are currently hundreds of gophers in existence, making it one of the primary resource discovery tools in use on the Internet today. One component of gopher is Veronica, which allows keyword searching of files and directories of all registered gophers around the world.

World Wide Web

World Wide Web, like gopher, provides simple access to a variety of Internet resources. In contrast to gopher, however, WWW uses a hypertext interface to navigate though documents distributed throughout the Internet.

Summary

The Internet is an exciting development for the library community because it represents a quantum leap in information access and professional communication. It is growing very rapidly, both in the number of information resources and services available, and in its global reach. It is a complex, and constantly evolving, global information system that has the potential to transform all sectors of society that utilize information. Libraries and information centres——because they are primary mediators of access to information——will be among those institutions that will be the most transformed.

Further reading

Comer, D.E. (1991). *Internetworking with TCP/IP. Volume 1*. (2nd Ed.). Englewood Cliffs: Prentice-Hall.

Krol E. (1992) *The whole Internet: user's guide and catalog*. Sebastopol, CA: O'Reilly and Associates.

Marine, A., Kirkpatrick, S., Neou, V., Ward, C. (1992). *Internet: getting started*. Menlo Park, CA: SRI International.

GLOBAL INFORMATION NETWORKS – HOW THEY WORK

by

Angelo Bodini
Projects and Engineering Department
European Space Agency, ESRIN
Telecommunications Unit
Via Galileo Galilei
00044 Frascati
Italy

Abstract

"Network" is a word largely used today in many different contexts, and its meaning is highly dependent on the context. This paper focuses on telecommunications networks. A first distinction is introduced between networks structures, services provided by networks, and services provided by Hosts connected to networks. Information Services belong normally to the last category. Connectivity is highlighted as a basic network service, to support interactive communications, end-to-end file transfers, and data dissemination.

Higher level network functions, such as electronic mail, conferences, bulletin boards and directories make use of lower level connectivity functions but provide services of a different nature, referred to as Value Added Services (VAS). Examples of this classification, which is mapped along the OSI reference model, are taken from Packet Switched Networks, Internet, Decnet, EARN/Bitnet and the protocols mostly used are outlined. A similar classification is to be made also for the gateways, i.e. devices and systems which interconnect different heterogeneous networks. The importance of gateways is invaluable, being the key to multiplying the connectivity of individual network implementations and fundamental steps towards the creation of a global virtual network for Information Services.

1. Introduction

It is not essential for users who need accessing, via networks, scientific and technical information to understand all technical mechanisms which implement the functionalities exhibited by telecommunication networks, in the same way as a driver does not need to know all the details about the engine of his car. Some hints however may be useful, especially when performances for given applications need to be tuned or when perspectives of new services are explored.

This paper focuses on some key aspects of networking and internetworking, in order to show which elements are activated inside a network when a key is pressed on a keyboard, and which concepts are hidden behind the often obscure jargon used by telecommunication specialists.

Despite what it might seem, there is nothing magic behind difficult words often used in the telecommunication field, although some concepts are indeed elaborate and equipment as well as procedures employed are in many cases at the front edge of the technology.

In approaching this area, as well as many other technical fields, a primary objective shall be the achievement of a clear and structured view of problems and concepts, and the understanding of which position each element of any complex system under consideration holds in the overall telecommunication architecture.

The architectural point of view will be the guideline of this discussion, in which some basic telecommunication concepts and techniques used today in the information networks will be reviewed, and mechanisms of real networks in operation will be explored, together with their interconnections.

2. Networks and Hosts

The first term deserving some clarification is "network". Many categories of systems are called networks, especially when they are composed of many elements coordinated towards a common goal, characterized by high level of mutual relations between their components, often dispersed in wide geographical areas. For example highways are referred to as networks, as well as railways, or a consortium of shipping companies or even spying organizations.

In the context of this paper the meaning of the term network is limited to those systems dealing with the transport and delivery of information by electronic means. In this sense we are really talking about telecommunication networks, although many important features are indeed common to other categories of networks, like for example the geographical dispersion

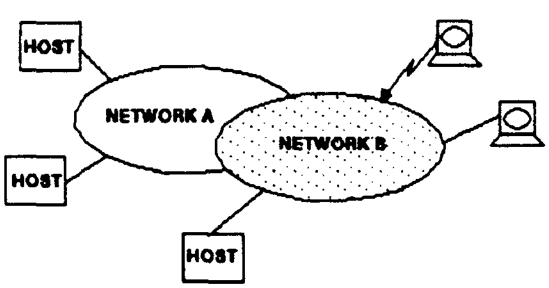
Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

of the entities involved, the presence of a technical and organizational structure, the existence of well defined working rules.

As telecommunication networks deal with the transport of information, in principle they are not involved in the production nor in the utilization of the information they intend to transport. Systems generating, absorbing or storing information are commonly defined as Hosts. Hosts are not structural components of a telecommunication network, but rather they use networks services to exchange information with other hosts or deliver it to users.

When a user is in front of a screen and makes use of an application, being an information retrieval system or an electronic catalogue of products, he should be aware that the functionalities specific to the applications are usually provided by the hosts running that application, while networks functions are confined to the transportation of data back and forth between his terminal and the remote host.

NETWORKS AND HOSTS



The situation is not always straightforward, as some applications presently in operation are in fact distributed among several hosts, which intercommunicate and exchange in real time queries and answers, while user sessions take place, although possibly presenting a uniform interface to end users. Hosts associated in this manner are also often referred to as being part of a network of hosts. This definition however is not consistent with the one adopted in the telecommunication area, and attention has to be made so as to not mix the concepts. The functionality associated with the provision and processing of information, typically implemented by hosts, should be regarded as something different from the functionality of transporting information, normally provided by networks, although both concur to the achievement of the overall result as it is perceived by the end user.

The relations between the above mentioned functionalities follow a general pattern where some systems (networks), which are themselves composed of subsystems (network components), provide services to other systems (hosts). These systems build on services received in order to provide more valuable services up to the final "System" which is the end user.

Any functional chain involved in remotely accessing information can be regarded as a structure, composed by several systems or components, each one making use of its internal resources and of a set of services, provided by other elements of the structure, in order to provide services at a higher level, in a sort of hierarchy which constitutes the architecture of the overall chain provided.

One of the primary merits of such a structured approach applied to telecommunication systems is to simplify dramatically the way in which they can be analyzed, described and understood, with the consequence of allowing difficult and complex problems to be broken down into more manageable ones.

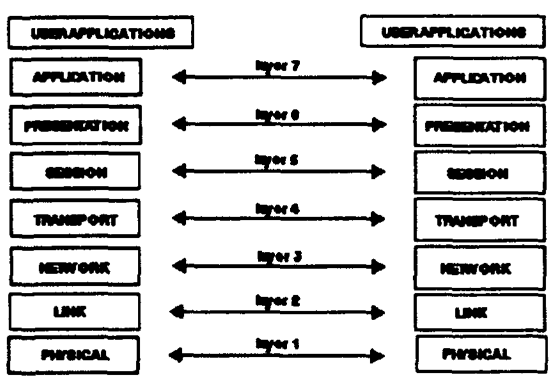
3. OSI Reference Model

The definition of a general functional structure to describe telecommunication systems was a topic of discussion for several years in the whole telecommunication world, until in the early 1980s the International Organization for Standardization (ISO) achieved a large consensus around a single network model, mainly under the pressure of equipment manufacturers interested in creating the environment for interoperable network implementations.

The consensus materialized in the definition of the Open System Interconnection (OSI) Reference Model, which was released as the international standard in 1984. It quickly became the primary architectural model for computer communications via telecommunication networks and the best tool to discuss the subject.

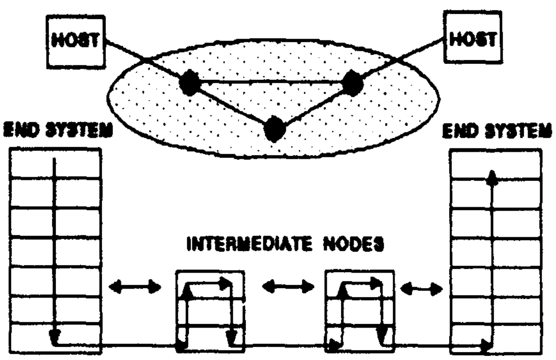
The OSI reference model subdivides the general problem of moving information between computers over a network into seven smaller and more manageable problems, and assigns each problem area to a specific portion of the model, called a layer.

OSI REFERENCE MODEL



The seven layers of the OSI reference model are structured in a hierarchy, so that each layer provides a service to the one above it, by using services provided by layers underneath, with the addition of new functions based on its internal resources. In this way, each layer adds some value to the services received when it passes in turn them to the layer immediately above.

INFORMATION FLOW ACROSS NETWORKS



In principle all seven layers are implemented in the remote systems wishing to communicate, and the OSI reference model describes how information makes its way from application programs in a system, through a network medium, to another application program in another computer. The way through is to descend from the application program through all layers in the transmitting system, to flow along the physical medium up to the receiving system and to raise again in the receiving system up to the target application.

Each layer in one system has to follow well-defined rules in order to be understood by its peer layer in another system. The set of such rules is called "protocol". Every protocol, when is coherent with the OSI reference model, is associated with one specific layer and addresses the specific problems that that layer is dealing with.

The three layers at the bottom address the general problem of connectivity and are identified as follows:

The Physical Layer defines the electrical, mechanical, procedural, functional specifications for activating, maintaining, deactivating the physical link between interconnected systems, e.g. voltage levels, clocks rates, physical connectors.

The Data Link Layer ensures reliable transfer of data across a physical link, and defines physical addressing, access control to the medium, error detection, recovery, flow control.

The Network Layer provides the path selection between two end systems which may be located in different networks, and rules to route information from a source to a destination across possibly multiple intermediate systems.

While the three lower layers are present in most of the equipment which can be found in a network, the four upper ones are related to problems in charge of the end systems, i.e. hosts, which make use of telecommunication networks.

The Transport layer has the purpose of providing a reliable transport service to the upper layers, by defining mechanisms for the establishment and disconnection of end-to-end logical links, fault recovery and flow control. By shielding all transport implementation details from the application, it allows the same transport service to be offered over a variety of different network media, and therefore can protect the application program against the change of network techniques.

The Session Layer organizes the dialogue between remote applications and manages their data exchange, for example by assigning the turn when to send or receive data.

The Presentation Layer addresses the problems of data format and agreements between different data representation in the end systems.

The Application Layer identifies and establishes the availability of intended communication partners, synchronizes cooperating applications and establishes agreements on procedures for control of data integrity, error recovery, overall session recovery.

The Application Layer, with all its services, is a component of a telecommunication architecture and is not to be confused with the computer programs implementing the information processing functions which the end user may require, like the construction of pictures or the retrieval of information. These programs are also called applications, hence some unavoidable confusion may take place. Application programs use the underlying telecommunication structure, including often the Application Layer of such structure, but are not part of it. Unfortunately the limited dictionary adopted and some lack of discipline in the use of common words does not help in enforcing a clear distinction of concepts.

4. Services provided by networks

The fundamental category of services provided by networks, which corresponds to their traditional "raison d'être", includes those services related to the provision of connectivity and the actual transport of information between remote systems.

The problem of transporting information is handled within the three lower layers of the OSI reference model, Physical, Data Link, Network. In the presence of a network providing connectivity services, the computer applications running in the hosts at the boundary of the network do not necessarily need to implement protocols at all seven layers foreseen by the model, but they can directly access the transport or the network layer, depending on the specific protocol suite selected.

Connectivity and information transport are the network services required to support Remote Login functions, interactive access from end users terminals to applications on remote hosts, data exchanges between cooperating applications implementing a client-server relationship.

Depending on the physical extension of the geographical area concerned with the connectivity services provided, networks are classified as Local Area Networks (LAN), Metropolitan Area Networks (MAN), or Wide Area Networks (WAN), and the technologies involved, the protocols adopted and the performances achieved are widely different in the three cases.

When protocols belonging to the higher layers of the OSI reference model are activated, networks can provide a wider set of services of a more sophisticated nature, called Value Added Services (VAS). They include File Transfer, Electronic Mail, Directory, Bulletin Boards, Conferences, Newsletters, etc.

The File Transfer service allows the access to file directories in remote hosts and the exchange of files on an end-to-end basis, according to one-to-one communication scheme.

Electronic Mail services allow the exchange of information in a store-and-forward fashion, according to one-to-one or one-to-many communication schemes. This service has achieved particular importance as most network providers around the world are currently offering it. Common characteristics to all mail services is the capability to accept, handle, route and deliver messages. Messages are composed of a "body" containing the text of the message, which in most systems may consist of a real text, but also of a binary file, a fax image, or a piece of voice or video, and an "envelope", which is the container of the body and carries information required for mail delivery, such as the address of the sender and the recipient, subject line, and requests for auxiliary services, like delivery notification or priority. The relation between the body and the envelope, format, submission rules, routing are all elements defined by the specific family of electronic mail protocols adopted by each electronic mail provider.

The Directory Services allow users to identify people they want to communicate with easily, by providing their electronic mail address, postal coordinates, etc., while the electronic Bulletin Boards, Conferences, Newsletters provide groups of users with the possibility to collect and share information of common interest; this kind of communication can be considered "group communication" and follows a many-to-many(from many originators to many recipients) communication scheme.

The provision of any of the Value Added Services mentioned before, requires always the availability of a telecommunication infrastructure implementing the lower layers services, usually Transport or Network. VAS offer to users a range of services of higher quality and more sophisticated than the pure transport of information. It is important to appreciate this difference in order to assess the nature of the services provided by different networks, and furthermore their capabilities. Often the same commercial names are used for different categories of services and hence misunderstandings may arise.

Network Management includes another range of services provided via protocols belonging to all seven layers of the OSI reference model. It is not a service designed for the large user community, but rather for a limited number of sophisticated network users or operators, who need to keep under control and to manage components and resources of the networks they expect services from. Network Management permits the accessing of information regarding the actual status of any equipment which is part of a telecommunication system, to receive alarms in real-time and to perform corrective interventions or reconfigurations actively. Functions of this kind are particularly relevant for service providers and network operators, and are not normally accessible to end users.

5. Information Networks

It is worth reviewing what are the network families most widely used today are in the technical information area, what services they are providing, and what networks can provide, or concur in providing a support to information services on a global basis. Attention will be focused on the European situation and the interconnection between Europe and the rest of the world.

The point of view of a large information service provider, such as the European Space Agency, is adopted hereby, especially in relation to the activities performed at the ESRIN establishment in Frascati, Italy. There, a range of services are put at the disposal of the scientific and technical community, including extensive archives and catalogues of Earth Remote Sensing data and images, order desk and scheduling information for acquiring data from the ERS-1 satellite, the European Space Information System for accessing astronomy and space science data, and the Information Retrieval

Service for scientific and technical bibliographical references.

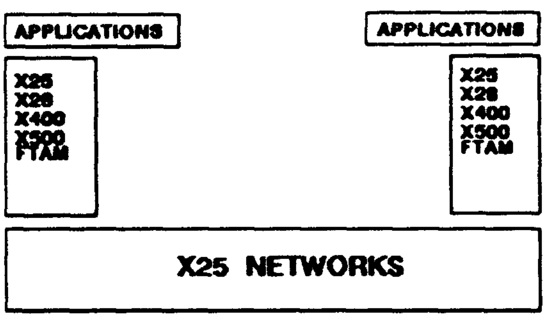
The primary problem of information services providers is via which networks can their end users access the services provided, while the symmetrical problem, seen from an end user point of view, is whether the services relevant for his activity are available via the network he is usually connected to, or not, and in this case how is he to get hold of them.

Only knowledge of the present network situation and of the trends of its evolution can help in deciding on practical steps to undertake and on strategic approaches to consider.

5.1 ISO-OSI based Public Networks

Europe is characterized by the widespread adoption of packet switching technology, which has led to the creation and the wide diffusion of packet switched networks in each European country. The set-up of national networks was encouraged by the initiative of the European Community in the late 1970s, to create a first international network nucleus, Euronet, and by all PTTs having adopted one common standard network access protocol, defined by the recommendation X25 of the Consultative Committee for International Telegraph and Telephone (CCITT) of the International Telecommunications Union (ITU). This protocol was subsequently adopted by the International Standard Organization (ISO) as the basis for the definition of the Network Layer protocol, e.g. Connectionless Network Protocol (CLNP).

OSI CONNECTIVITY VIA X25



The X.25 recommendation describes protocols related to the three bottom layers of the OSI reference model, indicating how to connect a packet mode terminal to a packet switched network, how to cut information in packets and how to submit it to a network for delivery to a correspondent. Routing, flow control, error detection and recovery are all problems addressed by the X25 recommendation, which was also designed to cope with the poor quality of the data circuits available at that time.

In each European country packet switching public data networks (PSPDNs) are operated by national PTTs or by recognized private operators, and despite the variety of names, geographically oriented like Itapac, Iberpac, Eirpac, Hellespac, Luxpac, or inspired by the technology, like Telepac, Datapak, Packet Switching Service, and so on, they basically provide the same services, i.e. transport and routing of information up to the Network layer. This service is offered to packet oriented terminals accessing via the X25 protocol, as well as to character oriented terminals accessing in asynchronous mode special network interfaces, called Packet Assembly and Disassembly (PAD) via a protocol defined by the CCITT recommendation X28.

In Europe the combination of public packet switched networks constitute today the most widespread and supported network infrastructure in operation. They are interconnected between them and with similar networks all over the world, for a total of more than 100 countries interconnected. The European Space Agency largely relies on interoperation with PSPDNs for the functionalities of its Data Dissemination Network.

Performances vary considerably from country to country, and make X25 networks best suited for applications designed around the use of character oriented terminals. The speed of most circuits employed varies from 9.6 kb/s to 64 kb/s, while in some countries links at 2 Mb/s are used for trunk connections. Some highly demanding applications, like those requiring intensive client-server data exchanges, or fast data transfer in large amounts cannot run comfortably on present X25 networks, and are also discouraged by the volume sensitive tariff structure generally adopted.

No public network offers a File Transfer Service as such, but hosts implementing ISO protocols operate easily across public networks, with the File Transfer Access and Management (FTAM) protocol which relies on X.25 for service provision at network level.

Public operators are also very active in the Value Added Services area and mainly as far as the provision of Electronic Mail is concerned. Practically all PTT Administrations in Europe have adopted the CCITT recommendation X400 for Message Handling Systems and are offering mail service based on it. X400 is a particularly powerful and complete protocol which defines the framework for services of store-and-forward nature, including Electronic Mail. Some initial fears about its complexity and implementation cost have now been dissipated by its widespread use and commercial availability.

As with the transport networks, also Mail services in Europe are organized on a national basis. Administration Management Domains are strictly organized with respect to national boundaries and the allocation of X400 addresses is coherent with this principle. PTTs provide mail connectivity on a worldwide basis with almost all public services in the world adopting the X400 approach.

5.2 ISO-OSI based Academic and Research Networks

The Academic environment in Europe has always been quite active in providing telecommunication facilities to the affiliated institutes, in parallel with the public Networks provided by PTT Administrations.

The European Community sponsored the set-up of a X25 network, called IXI (International X25 Infrastructure), as first nucleus of a network expected to become the backbone network for the whole Scientific and Research community, in the framework of the support to the European research.

Aimed at the provision of a packet switched data communication service for non commercial research applications, with an initial access speed of 64 kb/s, IXI played a pilot role in the interconnection of National Academic networks, although its popularity was superseded by other developments, namely the diffusion of Internet in Europe, which enjoyed a higher acceptance on the user side. IXI recently evolved towards a new structure and became the X25 component of a multiprotocol network, Europanet, operating with 2 Mb/s circuits, and is now potentially able to play a new significant role within the research community.

The most significant OSI related developments are taking place in the Value Added Services field, where the EEC has launched a family of projects (Y-NET, PARADISE, etc.), with the purpose of establishing a pan-European distributed OSI network to provide all participants in EEC research programs with improved communications and exchange of data, using OSI conformant software and equipment. The OSI-based network services include X400 Electronic Mail, X500 based Directories, and FTAM file transfer around Europe, for actual use by the Academic Institutions, and to pave the way for future public services to be offered.

Mail systems belonging to Y-NET, although based on X400, are not generally interconnected with those provided by public operators, with the exception of some remarkable cases in specific countries. However the same transport networks, i.e. public networks and Europanet, are generally used below.

The PARADISE project includes the implementation of Directory Services, oriented towards the provision of a world wide User Directory, a sort of general "white pages" service, with a distributed management arrangement, where each participating country, Company or Institute retains independence and the full responsibility of creating, maintaining and updating their member list. The standard protocols adopted (X500) define how to access, research and exchange data with remote directories and how to make a directory remotely accessible by all participants.

5.3 DECNET based networks

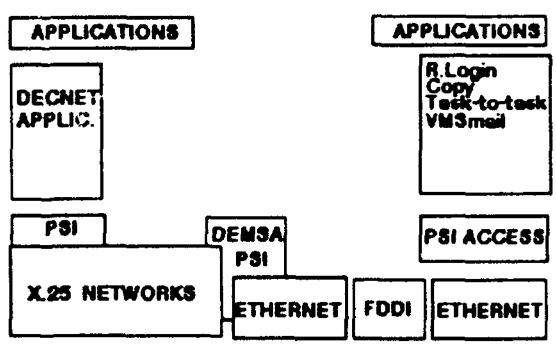
Decnet is a collection of protocols, based on digital Equipment Corporation (DEC) products, which have been adopted by a number of networks in the world, connecting mainly computer centers and scientific institutes involved in High Energy Physics and in Space Physics. Core networks initially included the Space Physics Analysis Network in USA (US-SPAN), the European Space Physics Analysis Network (E-SPAN), the European High Energy Physics Network (E-HEPNET) and the DECNET portion of the Energy Science Network, its counterpart in the USA.

All these networks, and the many other coordinated to them share a common address space, coordinate the allocation of address areas, cooperate in the management of routing rules and provide a general connectivity among them, although developed largely independently of each other.

The common characteristics of these networks are that they are not for public use, but in principle support only institutions involved in their specific discipline.

Decnet includes protocols defined at all the seven layers of the OSI reference model, which are suitable for LAN and WAN network technologies. In particular, from the connectivity point of view, remote Login and task-to-task communications are supported across ethernet or FDDI LANs as well as across geographical networks. Decnet transport layer in addition is capable of using network services provided by X.25 networks, and this facility made possible for many users in Europe to join the SPAN or HEPNET communities, by exploiting the connectivity provided by the public X.25 infrastructure, and for SPAN and HEPNET users to access external services, available on public X25 networks.

DECNET CONNECTIVITY VIA X25



The most significant trend today is for DECNET protocols to run over connectivity services available via Internet, with the consequence that from the point of view of the geographical distribution and the share of circuits capacity the two networks tend to overlap and to become indistinguishable, DECNET being operated, when required, almost as an end to end protocol across Internet.

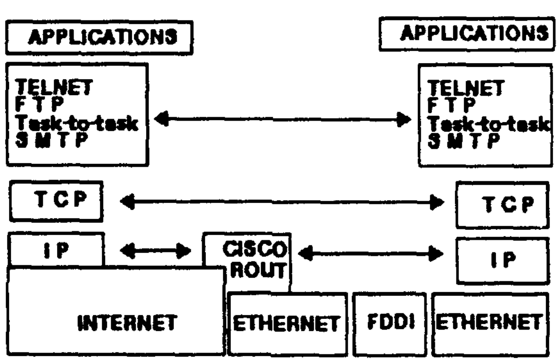
Among the Value Added Services, in the upper layers of the OSI reference model, DECNET provides a File Transfer Service, activated by the command "Copy" and Electronic Mail services, of which VMSmail is the most popular. These services are available and widely used by all hosts and users connected to SPAN and HEPNET.

DECNET is currently undertaking a global migration from the present set of protocols, defined as Phase 4, towards a new family of protocols, known as Decnet Phase 5, with the aim of becoming fully compliant, and therefore capable of interoperation, with ISO OSI protocol implementations. This is a challenging move and, because of the large number of installations existing in the world, is expected to last several years. During the transition period the two DECNET phases will coexist, horizontally crossing the entire network family.

5.4 Internet

Internet is a loosely organized system of interconnected computer networks, which primarily serves the research and educational community. All interconnected networks have the common characteristic of making use of the family of protocols identified by the most common ones, the Transmission Control Protocol (TCP) and the Internet Protocol (IP), and sharing a common address space. When capital "I" is used, Internet refers to the actual TCP/IP based worldwide network system, while with lower case "i" the term is used for any set of networks using the same suite (TCP/IP) of protocols, thus ensuring interoperability, but not necessarily interconnected to the world-wide system.

TCP/IP CONNECTIVITY VIA INTERNET



The first nucleus of Internet was ARPANET, started in 1969, subsequently adopted by the Defense Advanced Research Project Agency (DARPA) which very early on understood the potential of packet switching technology in solving the problem of communicating between heterogeneous computer systems. DARPA funded the development of a series of protocols, equally well-suited for Local Area Network as well as for Wide Area Network Communications, and made them freely available to the users community.

In the recent years the TCP/IP protocol suite enjoyed the benefit of being included by most vendors in their UNIX operating systems offering, which led to an unexpected diffusion, parallel to the success of UNIX, in most academic and research environments.

The reason for the widespread enthusiasm for Internet is not that it is the optimal high speed data network. In fact the networks main family of protocols, TCP/IP, is now quite old and less efficient than newer approaches to high speed data networking, like Frame Relay or SMDS (Switched Multimegabit Data Service). Its success stems from the fact that Internet is often today the only possible outlet for eager users which offers a standardized and stable interface, along with a deliberate focus on openness and interconnection. These features make Internet extremely attractive to very different groups of users in corporations, government agencies and academic institutions.

Today Internet is a logical network connecting more than 2000 networks all over the world. In USA regional networks have grown around the National Science Foundation initiative NSFNET, which acts as a federal backbone and coordinate the access of a wide number of metropolitan or campus wide networks. Via two Federal Interchange points the NSFNET is connected to the other national backbones, like the NASA Science Internet (NSI), which is dedicated to the support of space science and NASA related activities, and to international networks.

In Europe, Academic communities in most Western European countries have set-up TCP/IP networks on a national basis, and made them part of Internet. The most relevant implementations include INFN and GARR in Italy, WIN in Germany, Janet in UK, RENATER in France, NORDUNET in Sweden, Norway, Denmark and Finland, SURFNET in The Netherlands, ACONET in Austria, etc.

The network access policy, and the attribution of new addresses, etc. are considered as national matters and are governed in each country by the research institute in charge of the network organization. In particular the possible access extension to commercial entities is considered a national concern, in accordance with general guidelines indicating the acceptable use of the network.

Infra-European interconnections between different countries are mainly arranged on a bilateral basis, with a few organizations, acting as focal connection points for national networks. A typical example is CERN, the European Organization for Nuclear Research, which in addition to being the natural termination point for all traffic-related to High Energy Physics, which counts for the vast majority of scientific traffic exchanged across Europe, also acts as switching point for a number of national networks between them and with Internet in USA.

Specific networks were created in the past with the aim of performing this very function, e.g. E-BONE, but the emerging initiative in this direction is now represented by Europanet, a multiprotocol network created to satisfy the international networking requirements of the European research community.

Another important implementation is being performed by the European Space Agency with the creation of ESINET, a TCP/IP based network extended to ESA establishments in Europe and connected to NSI. ESINET is a discipline-oriented network, as it only supports communications related to ESA activities.

In parallel to the academic networks, commercially oriented organizations are managing the access to Internet on behalf of commercial users. An example is EUNET, which has set up an independent telecommunication structure in Europe, with Points of Presence in all European countries.

Connection to Internet requires the assignment of an Internet IP network number, unique in the world, to be obtained by the Network Information Center (NIC), which acts as a global coordination entity.

A general tendency to progressively adopt TCP/IP connectivity and migrate towards Internet, from other networks, mainly DECNET, can be clearly identified in Europe. This is due to a combination of factors, such as the growing acceptance of UNIX as the basic operating system for many scientific applications, the establishment, as a consequence, of a multivendor working environment, and the increased quality of data circuits, which in addition allow good performance to protocols inherently unreliable, such as IP.

IP is a simple routing protocol, providing an inter network service, without error detection or recovery. These are left to the TCP protocol, which in addition to breaking data into datagrams and reassembling at the other end, handles error recovery, by resending lost or garbled datagrams, and controls data flow. The higher layers protocols providing remote Login and interactive access, such as Telnet, and task-to-tasks specific application protocols, are based on TCP, which belongs to the transport layer.

The File Transfer Protocol (FTP), which allows the transfer of files between two hosts, at the application layer, is also based on TCP/IP. It is possible to connect to a remote host, via FTP, to provide an account identifier and a password, and to "get" or "put" files from and to the host. Many hosts in Internet use this protocol to distribute data, programs and documents which are publicly available, and call this service "anonymous FTP" because of the use of the word "anonymous" as generic account identifier.

The TCP/IP protocol suite includes an electronic mail protocol called Simple Mail Transport Protocol (SMTP), which standardizes the exchange of mail between any hosts in the Internet. It is usually known as Internet mail and is available to any piece of equipment capable of being connected to Internet. The Internet mail is probably the most widely used mailing system in the world today.

A variety of additional protocols are available and operated on Internet, despite at present being less popular than those mentioned above, while new protocols are continuously experimented and proposed, addressing new functions, such as data broadcasting, transport of voice, or videoconferencing.

The determination to remain "informal", in the organization, operation and in the management of evolution, is an important peculiarity of the Internet community, originating from the strong presence of Universities and Academic institutions. For example protocol definitions are called Request For Comments (RFC), indicating that when no further comments are collected in relation to a given protocol, it becomes "de facto" accepted as a standard. Such a process, although it does not guarantee high stability to protocols, has nevertheless proved to have the merit of achieving convergence and results much faster than the formal processes adopted within the official standard organizations (e.g. ISO, CCITT, ETSI).

5.5 BITNET/EARN

BITNET (Because It's Time Network) started in 1981 as a small, cooperative store-and-forward network of IBM computers centered around the City University of New York. Today it is a worldwide network, linking institutional and departmental computers, including universities, colleges, and collaborating research centers. The European proliferation of BITNET constituted the European Academic Research Network (EARN). Both networks have as their purposes the facilitation of the exchange of non commercial information, consistent with the academic aims of its members.

Hosts and nodes on BITNET communicate with each other using software implementing IBM's Network Job Entry/Network Job Interface (NJE/NJI) protocol. File Transfer and Electronic Mail services are essentially available via this protocol, while no interactive application can be supported.

The importance of BITNET in Europe tends to rapidly decrease in favor of Internet, as soon as TCP/IP protocols become available on the various hosts concerned. However the very wide diffusion that it had in the world, including the developing countries, and its permanence in many institutes, often makes BITNET the only viable way of exchanging information in electronic form.

6 Convergence and coexistence

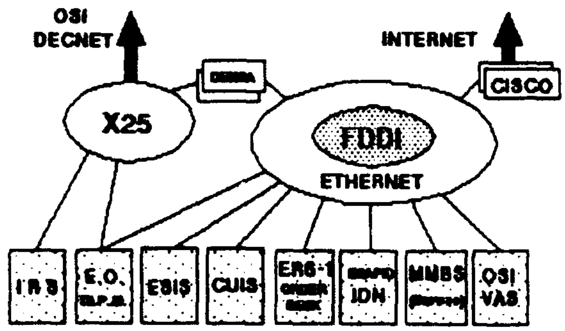
As previously mentioned, the biggest problem faced by information systems users facing such a variety of networks is which one to use, or where to connect to. Obviously there is no straightforward answer valid for everybody. While the main consideration should be related to the reacheability of the intended correspondents or the required service hosts, often practical circumstances make the alternatives for a real choice very limited or almost inexistent. The ideal situation is for a user to be connected only to one network and still be able to reach all hosts relevant to his activities, independently from which network they belong to. This ideal would be for all categories of services desired, whether pure connectivity, mail, or other Value Added Services. The obvious example taken is the international public telephone network, where such wide connectivity is today a reality.

The basic difficulty arises from the fact that the different protocol suites are in general incompatible and unable to interoperate. So OSI Hosts cannot establish a working session with a TCP/IP correspondent, or exchange data with a DECNET system.

Any hope for a general migration towards a unique family of protocols, which has always been the dream of all standardization bodies, is certainly unrealistic, at least for the near future. In absence of such a convergence the policy of the most active Service Providers is to become themselves multistandard, i.e. to operate on the same machine several telecommunication protocol suites in parallel, in order to become compatible with any remote host or terminal. The implication is that Applications have to be submitted to adaptations, often non trivial, as far as their interfaces to communication services are concerned.

ESA Data Dissemination Network

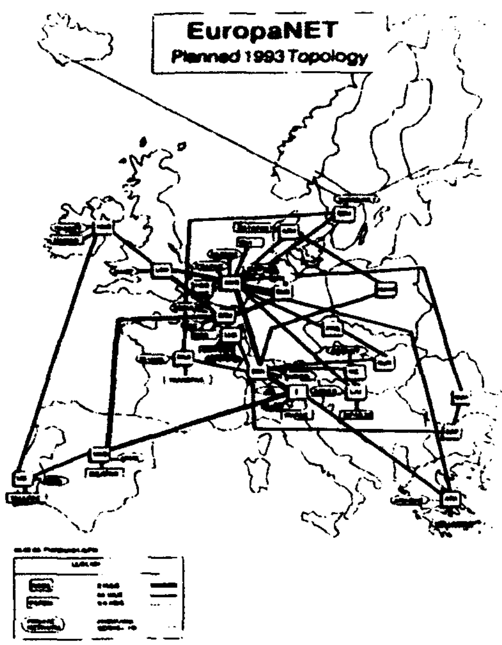
LOCAL CONNECTIVITY IN ESRIN



This approach is largely adopted within the European Space Agency, where all Information Services provided by the ESRIN establishment are gradually becoming equally accessible by OSI, Internet or DECNET remote users. Physical connections to the networks are however different for the different protocols, as with, at present, the infrastructures carrying TCP/IP traffic from those transporting OSI or DECNET.

On telecommunication networks side, however, a tendency is arising to create multiprotocol networks in Europe, i.e. networking structures where several communication protocol families can coexist and be supported concurrently by the same circuits and interfaces. The reason for this is, among others, the continued high cost of data circuits, which discourages the creation of separate infrastructures, specialized for protocol. A multiprotocol network is also particularly useful for those users who want to migrate smoothly from one protocol to another one, as they are also not forced to change, in addition to the protocols, the network they are connected to.

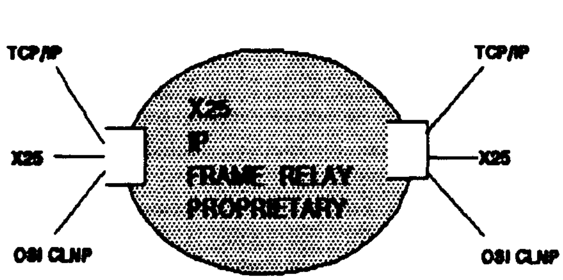
One of the most interesting initiatives, started under the auspices of the European Community, is Europanet. The Europanet Service is a virtual private network service delivered over a Multiprotocol Backbone infrastructure. The Service is designed and implemented to respond to the needs of the European Research Community, with support, as access protocols, of TCP/IP, ISO CLNP and X25 packet switching as well as Frame Relay. The Europanet Service is actually being provided to the European Research community, including Research Departments of Industrial Enterprises, and Research and Development Programs of Governments and the Commission of the European Communities, by Unisource, a consortium of Dutch, Swedish and Swiss PTTs, under the management of the British company DANTE Ltd. Shareholders of DANTE (the name stands for Delivery of Advanced Network Technology to Europe) are the main national research organizations in Europe, already participating as Full National Members in the European Research Network Association (RARE).



The Europanet Service can be accessed at Points of Presence in all European countries, either directly or through connected national research networks, and has connection to similar services worldwide.

The same approach is being currently adopted by private companies or PTT consortia, which have the aim of playing a dominant role in the handling of international data traffic in Europe in the future, or of becoming a telecommunication carrier on a global basis. A proliferation of multiprotocol networks is expected in the forthcoming years, although it is difficult to predict which one will survive the unavoidable competition which will ensue.

CONVERGENCE : MULTIPROTOCOL NETWORKS

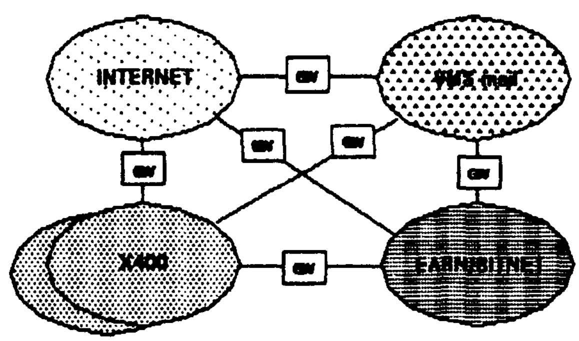


From a technical point of view a multiprotocol network faces the choice of which protocol shall be adopted as fundamental carrier, on which to encapsulate all the other protocols to transport. While Europanet selected a proprietary protocol as carrier, most commercial operators are oriented towards the adoption of Frame Relay, a protocol designed along the conceptual line of X25, but very much simplified and tuned for the use of the high quality digital circuits available today.

A general tendency is also emerging to accelerate the utilization of high speed links, the adoption of ATM switching technology and the integration in one single broadband network of voice, video and data traffic, while maintaining the differences in the accessing protocols. The multiplicity of protocols present today will be first of all confined to the boundary of the network, and will furthermore be greatly reduced, as two protocol families will probably remain, TCP/IP and OSI. In particular at the higher levels of the OSI reference model a clear convergence is taking place, oriented towards the adoption of OSI protocols, but running on different network and transport platforms. In this sense interoperability issues will continue to keep network specialists busy, although against a more limited number of viable alternatives.

The situation is substantially more comfortable in the field of Electronic Mail systems. A number of gateways are in operation in the world in order to convert the format and to route mail messages from one system to another. The fact however that not all mailing systems support the same set of auxiliary services, e.g. the delivery notification, cannot be overcome by gateways, and this leaves users of different networks with inherently uneven performances. The case of user interfaces, which remain specific to each system and have no impact on the overall interoperability, is instead different and not relevant for interoperability.

CONVERGENCE : GATEWAYS for MAIL



Technically Gateways can operate from any mailing system to any other, and within Scientific and Academic communities electronic mail is no longer a problem. However as soon as users on public mailing systems need to be involved in the exchange, regulatory barriers start appearing, not necessarily because of the high costs involved, but because of the absence of a general

convention between systems based on fundamentally different funding assumptions and therefore charging principles. Commercial public systems generally tend to apply tariffs proportional to the use, while Academic networks are oriented towards flat rate approaches. This is another element of inconsistency that must be overcome before the world can be really considered as a unique interconnected system.

References

Cisco Systems, Internetworking Technology Overview, Corporate Headquarters, Menlo Park, California, USA, 1993

NSI Sterling Software, The NASA Science Internet, A Summary,NASA Science Internet Office, AMES Research Center, Moffet Field, California, USA, August 1992

Hart J.H., Reed R.R., Bar F., The Building of Internet, Telecommunication Policy, Volume 16, Number 8, November 1992, Butterworth-Heinemann Ltd., Linacre House, Jordan Hill, Oxford, 1992

La Quay, Tracy L., Directory of Computer Networks, *Digital Press/Prentice-Hall*, USA, 1990

THE CANADIAN NETWORK FOR THE ADVANCEMENT OF RESEARCH, INDUSTRY & EDUCATION (CANARIE)

by

W.J. Padden
Industry and Science Canada
Information Technologies
Industries Branch
235 Queen Street
Ottawa, Ontario K1A 0H5
Canada

OVERVIEW

The extension of voice communications to include the simultaneous telecommunication of data, text and imagery in combined multimedia formats provides researchers with a radically new means of engaging in collaborative research and development with one another. Being able to simultaneously share and manipulate the same images and data on a given topic appearing on their computer screens, at any distance from one another, while at the same time being able to converse on that topic, promises more innovative, more efficient, and less time-consuming ways of conducting research and development.

Whereas the speed of computers has risen exponentially over the past two decades, the capacity of telephone lines to carry the data that computers are capable of generating has not kept pace, albeit for very justifiable reasons associated with the traditional environment of the telephone industry. Consequently, governments around the world are fostering the development of very high speed transmission facilities and international telecommunications protocols, as exemplified by the Open Systems Interconnection (OSI) models, in order to exploit the social and economic benefits of multimedia networking in business, research and education.

CANARIE is a proposed, national, high-speed backbone network which, when fully implemented in a multimedia format, will enable engineers and research scientists in any field, and in laboratories anywhere in Canada, to collaborate with one another and with their international colleagues on research and development projects of common interest, without the inhibitions presently imposed by the lack of proximity to one another. It may be anticipated that this "shortening of economic distance" among them will lead to an increase in research and development activity and improvements in their productivity. It will also tend to accelerate the diffusion of technology from the laboratories into marketable products, and serve as a model for improving performance, hence competitiveness, in every sector of the economy.

The intent of government sponsorship of CANARIE is to encourage the development of such a network before it would otherwise become commercially available in response to the normal evolution of a general market for broadband services. From this perspective CANARIE will serve both as a catalyst and as a vehicle for the development of future generations of network technologies, products, applications and services within Canada's Information Technologies sector. This development activity will focus primarily on lightwave technologies, with reference to standards-based open systems in general, and to the OSI Model in particular. It will encourage Canadian information technology companies to participate in the provisioning of CANARIE as it evolves, and enhance their ability to compete in emerging international markets for related multimedia networking technologies.

CANARIE will also provide a test medium which firms in other sectors and organisations, as well as the information technology industry, may use to test newly developed technologies, networking tools and applications.

CANARIE will facilitate the use and further development of national databases and access to remote facilities such as supercomputers. In this way it will complement research and development activities, provide training opportunities for research scientists, engineers and students, and serve as a model for distance learning.

CANARIE is a response to government-sponsored network initiatives in other countries and will not compete in any way with commercial network services.

THE CANARIE PROGRAM

A Business Plan, representing the views of industry, the universities, the regional networks and the provincial governments, was submitted to Industry & Science Canada in December, 1992. It recommended that CANARIE be developed in three phases over a period of seven years. Phase 1 is scheduled for completion by 31 March, 1995 and includes the following elements:

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

1. An upgrade of the existing, operational, research and development backbone network, known as CA*net, which links Canada's ten regional networks. This will include an increase in the data transmission rate from 56 thousand bits per second (kilobits/sec) to 1.544 million bits per second (Megabits/sec) and the development of new applications which will make CA*net more attractive to a broader user community.

2. The establishment of an experimental test network by the carriers, Stentor and Unitel, including the provision of laboratory facilities at separate sites, and the linking of these sites with fibre optic cable. This network will enable the testing and showcasing of next-generation lightwave technologies at transmission rates of billions of bits per second (Gigabits/sec) and higher. (A 1 Gigabit/sec transmission line could transfer the Encyclopedia Britannica across Canada in about a second.)

3. The initiation of a research and development program to encourage the development and application of lightwave technologies, particularly in innovative, smaller to medium-sized firms which would be unable to assume the technical and marketing risks involved in the absence of funding assistance.

During Phase I, a number of issues will have to be examined, such as network access and affordability, the rationalization of conflicting standards, privacy and security, and emerging technologies in relation to the regulatory environment.

According to the Business Plan, Phase II would see the operational research and development network, CA*net, upgraded to transmission rates of 45 Mbits/sec or higher as lightwave technologies were deployed. The core test network would be expanded across Canada and finally, during Phase III, both the experimental and operational networks would merge to form CANARIE. At that time the network would be financially viable and further government support would be unnecessary. Eventually CANARIE would become indistinguishable from other networks, as the general market for commercial broadband services evolved.

PROGRAM FUNDING

The Business Plan estimates that the total cost of implementing a national fibre optic backbone would be some $985 million. The cost of upgrading the regional networks to achieve technological parity with the national backbone is an estimated minimum of $200 million, depending on timing and other variables, bringing the total cost of the CANARIE Program to approximately $1.2 billion. The recommended combined contribution by the federal government is $208 million.

The estimated cost of Phase 1 is $115 million, of which the private sector contribution will approximate $89 million in cash and in kind. The balance of $26 million will be provided in accordance with a Cabinet decision to support CANARIE, reached on February 25, 1993, and under terms and conditions subsequently approved by Treasury Board on March 17, 1993. Of the $26 million, Treasury Board has authorised an expenditure of $7.5 million in the current government fiscal year. The remaining $18.5 million will be expended in fiscal year 1994-95. Expenditures on elements 1, 2 and 3 above will be $7 million, $3 million and $16 million respectively. In addition, Treasury Board has authorised Operations & Maintenance expenditures of $500,000 in each fiscal year. These expenditures will include Program Monitoring activities to ensure compliance with the Financial Administration Act, Consulting Fees in support of a Business Plan for Phases 2 and 3, Program Evaluation, and Policy Consultations in the context of harmonisation with other international research network initiatives.

MANAGEMENT

The management and implementation of the CANARIE Program will be carried out by CANARIE Inc. under the terms and conditions of a Contribution Agreement with Industry & Science Canada, which took effect on June 21, 1993.

CANARIE Inc. is a not-for-profit organization which was incorporated on March 5, 1993. The board of directors is composed of industry executives and members of the academic and regional network communities, under the chairmanship of a senior partner with the consulting company Ernst and Young. The immediate task of the board of directors is the recruitment of staff to plan and execute Phase 1.

CANARIE Inc. will absorb its own administrative and overhead expenses through a levy on its members according to their means. Although membership is encouraged, it is not a condition of access to CANARIE Program funds. The latter will be assessed under qualifying criteria approved by Treasury Board and embodied in the Contribution Agreement.

DELIVERABLES

Under the terms of the Contribution Agreement, CANARIE Inc. will prepare an Operations Plan detailing its conduct of Phase 1. It will be submitted to Industry and Science Canada for approval within the next few weeks. The Plan and subsequent related progress reports throughout the course of Phase 1 will be used, among other factors, as the primary basis for payments of CANARIE Inc.'s claims against Program funds.

Towards the end of Phase 1, CANARIE Inc. will provide Industry & Science Canada with a detailed Business Plan which will determine the feasibility of continuing with the Program.

Standards for data and document interchange.

by

Geoffrey A. Stephenson

Knowledge Exchange Technologies Ltd.
11 Rue du Plebiscite
L-8037 Strassen
Luxembourg

INTRODUCTION

Background

Information flow is a driving force in modern society. The list of ways in which information flow is necessary for commercial and cultural activity is endless. We use information every day, whether it is to look up a telephone number, check how to programme the video recorder, or plan a holiday. Hundred of millions of newspapers and magazines are published every week. In large commercial organisations a significant part of the work force are employed solely in handling information; they deal with such things as orders, invoices, payments, distribution plans, advertising, and technical documentation. Many organisations exist with the single aim of producing and distributing information. Prime examples are publishers, TV companies, market research organisations, design teams, consultancies, financial advisers, universities, employment agencies and national statistical offices.

Traditionally information has been distributed on paper as text and pictures. Printing in Europe began in the period 1440-1450 and grew at an astonishing rate. One estimate puts the total of books printed by 1500 at 20 million, for a European population of about 70 million. In the last hundred years the invention of records, radio, film and television has progressively extended the forms in which information can be distributed to include sound and motion pictures.

Text and pictures printed on paper have the great advantage that they can be directly read by the recipient. No matter what technology has been used by the printer to produce the publication, the reader can see the result. The more recently developed delivery mechanisms require the use of machines to convert the information into a form recognisable to the viewer. The success of the cinema, gramophone and television has depended on the rapid introduction of public standards that enabled the makers of the machines required for the projection of information to build systems that would accept a standard input format. In turn the availability of such projection devices encouraged the production and distribution of information products (records, films, TV programmes) for a mass market.

The use of computers, in conjunction with the telephone network has introduced new possibilities for information interchange. Most people are now familiar with the facsimile machine or fax. A fax is simply a special purpose computer with digitising and printing capabilities attached to a telephone line. The phenomenal growth in the number of fax devices installed has revolutionised the transfer of small quantities of information in the last five years. The fact that fax takes information on paper as its input and provides output in the same form, is one of the reasons for its success. Information interchange works only if the recipient can read the message.

The growing availability of microcomputers and telecommunications networks, both in the office and the home, opens new opportunities for information distribution similar to those at the birth of radio and television. If these opportunities are to be seized then the encoding of information must be standardised. In the late 1980s the *de facto* acceptance of the IBM PC and its operating system as industry standards, opened the market for microcomputer software. The rapid growth in the market, that followed, was similar to that in the video market after the general acceptance of the VHS standard for video recorders, a decade earlier. If electronic information products are to be sold in the same way that print is at present, then a high level of standardisation is necessary.

Open information interchange

To a considerable degree, the international standards necessary for the interchange of electronic documents are either available or under development. Failure to apply the standards is a potential barrier to the rapid expansion of the electronic information market. Lack of awareness of the standards for encoding information and the small size of the market for electronic information products have resulted in dominance at present by incompatible proprietary solutions. The Commission of the European

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

Communities (CEC) plays an active role in promoting the use and development of standards throughout European industry. As part of the Information Market Policy Actions (IMPACT) programme of Directorate General XIII E, the CEC launched an initiative in 1991 to promote the use of encoding standards for information in an electronic form. This initiative is Open Information Interchange (OII) [1].

OII is deliberately limited in its scope to media independent coding standards. When an electronic document is transmitted over a network between computers, or stored on a magnetic tape or disk, a certain amount of electronic information is added that is purely concerned with the operation of the computers and other equipment involved. This information ensures that such aspects as synchronisation between the devices and data integrity are preserved. It is added by the originating equipment and stripped away by the receiver before the 'real' information is presented to the reader. Standards exist for this control information. In addition standards are required to ensure that the transfer media are physically and electronically compatible between different manufacturers (plugs and cables must be compatible, track layout on diskettes or tapes must be the same). OII is only concerned with the 'real' information and thus excludes the standards that are concerned with the specific transmission media, whether on-line (Open Systems Interconnection standards, OSI) or off-line (CD-ROM and other media standards such as High Sierra or ISO 9660). The value of standardising information products at such a machine independent level lies in the possibility of preserving the information beyond the life of the media. Technology changes rapidly and the life of a particular device is short. If information coding is tied to specific media then the information products will not outlive the contemporary technologies. If the coding is media independent, then the information can be transferred between different media and preserved through time. Printed information lasts for centuries and if it is to be replaced, even partly, by electronic products then they too must have a long life.

Information and industry in the 1990s

Recent economic studies [2.3] suggest that the intensive application of information technology (IT) in industry is changing the way leading companies are developing their commercial strategy. In the past IT systems were applied in companies to support existing corporate activities. Stock control, accounting systems, distribution and production were automated in the sense that existing manual systems were transferred to the computer, with improved accuracy, speed and flexibility in reporting. In the new scenarios, IT and corporate strategy are jointly developed to increase company competitivity. Examples of this approach can be found in most industrial sectors and in a variety of forms.

In the engineering industry, for example, it was seen that productivity improvements were possible if stock levels of components and finished product could be reduced. In order to achieve this the required quantities of materials and other parts need to be delivered to the manufacture or assembly site at the moment they are needed, and the finished product must be produced in line with customer delivery requirements. This can only happen if the complete production control system is integrated together with the order system, and control of sub-contractors. Such systems, known as 'just in time' manufacturing are possible if the IT and other systems are fully integrated. In addition, quality control has to be improved to ensure that the whole operation is not held up by faulty parts. Total quality control is a system, again based on intensive application of IT, that ensures that quality is maintained throughout the production process. These technologies are being extended by the application of computer integrated manufacture (CIM), robotics and mechatronics to a point where it is conceivable that the whole process of production from the acceptance of a design to the delivery of a finished product may become completely automated.

Areas of industrial activity that have been untouched previously by automation are being brought into the new integrated systems for quality control purposes. Technical documentation is an important example of this trend. Installation and maintenance manuals are in the main produced manually and their production is not properly connected into the product design system. As a result, modifications frequently are not reflected in the documentation and updates are not distributed in a timely way. The attention now being paid to quality control has revealed that significant amounts of downtime in industry are due to poor installation and maintenance of equipment, in turn due to out-of-date or inaccurate documentation. The solution to this problem is seen to lie in better control of the production and updating of documents through the application of computer methods. Some companies now recognise good documentation as an important selling point. The importance of such documentation systems is also highlighted in the new international standards for quality control (ISO 9000) that are being introduced into industry. A contributory problem with documentation is the increased complexity of the machines we use. For example, where a BMW 3.0 Si car in the 1970s might include in its assembly 4 electric motors, 8 control units and 27 switches, the comparable version in the 1980s had 73 electric motors, 25 control units and 60 switches [4]. The increase in complexity generates thicker manuals with a corresponding increased cost in document management.

In all these examples, the improvements in productivity, the competitive advantage developed depends on the improved use of information. Commentators have referred to the "information economy", peopled by "knowledge workers", the "computer integrated company" and the "networked" or "intelligent" enterprise. These buzzwords try to capture the essence of the evolving environment where information is counted as a tangible resource that is vital to the competitivity of the company and that therefore must be managed effectively. Companies that have adopted this new way of thinking and have increased their productivity, put commercial pressure on less advance companies forcing them to catch up or go out of business. This pressure is accentuated by the growth of global competition. Markets in individual countries are being progressively opened to competition from other countries through regional economic agreements, and thus, to competition from the most efficient producers world-wide. The results of such competition on weak companies can already be seen in traditional European industries such as ship building and steel production.

In the 1990s we shall see an increased awareness of the need to use information effectively, and pressure for information to be available, when and where it is needed, in an immediately useful form. Information has become an increasingly important input to the production of goods and services; information about the market, information about the production process, information about consumer needs. On the one hand, the ever growing complexity of the man-made environment, and on the other hand, the global market in which much of commerce now operates, can only accelerate these trends. Companies will insist that their suppliers provide them with information quickly and accurately, whether it is price lists, delivery dates, maintenance manuals or product specifications. Similar demands will come from consumers, and, driven by an increased concern for environmental and safety issues, from government regulators.

The information market itself will not be immune to these pressures. The workers in industry who are responsible for processing the information that improves productivity, are increasingly connected into corporate telecommunications networks that offer integrated information handling capabilities. Information from external sources, both associated companies and information vendors, will have to be made available in forms that can be easily imported into these corporate systems. Just as foreign exchange and securities dealers of today are dependent on sophisticated multi-screen workstations, with integrated feeds from financial information service providers, so their counterparts in publishing, product design, and management information services will need similar tools in the future to work effectively.

Ease of use

People prefer the things they know to the unknown, we are all risk averse. In practise, this means that people shy away from uncertainty. The available methods of information interchange (post, fax, e-mail, etc.) are judged against the criteria of cost effectiveness and perceived benefits, with uncertainty in mind. The questions asked are: can we make it work? will it reach the client? will he/she be able to read it? Ease of use and fidelity of reproduction have clearly been keys to success. In the mid 1980s the number of e-mail and fax users in Europe were in the low hundreds of thousands (less than half a million), by 1991 the e-mail users had grown to about 1 million whilst the fax users had grown to over 3 million. At the same time the telex users had stayed roughly constant just above the half million mark. The phenomenal growth in the fax market has its roots in the demand for fast accurate information supply identified above. Yet comparing fax with e-mail, the latter has far better potential. Fax provides only a paper copy, if the information is to be used in a computer it must be rekeyed (or converted to a revisable form with OCR software), the quality is poorer (compare fax output with a laser print), and the transmission costs are higher (image versus character string). User friendliness has been a major factor in the success of fax. Anyone who can operate a telephone and a photocopier can use a fax machine. The fax machine is sold as a unit and is standard throughout the world. E-mail requires a mixture of hardware and software (PC, modem, communications software, word processor) to be bought and installed, is difficult to use because it has no user friendly interface, and due to the lack of standards the software will not be compatible with the majority of callable sites and so a true copy of the document will not be transferable.

The problem we are faced with is how to combine the required integration of applications with the ease of use and interoperability that is the key to take up. Application integration is occurring at several levels. At one extreme we have the individual user, carrying out some task, who needs to access information and use it in his task-specific environment. The user may be an engineer importing material property data into a design, a foreign exchange dealer accessing the latest currency rates, or a manager incorporating data from a spreadsheet into a report or prices into a tender. At the other end of the scale, corporations need to integrate production and distribution systems, integrate material supply from sub-contractors into the production cycle, integrate technical documentation with product design. Information must be passed between applications running on the same local system, but also between systems belonging to independent organisations.

The general process developed to deal with this sort of problem is standardisation, whether standards determined by official standards bodies, or industry standards sanctioned by the market or a group of suppliers. We take many of the results of standardisation for granted. When we buy an electrical appliance, we expect that when we arrive home the plug will match the socket in our wall. Those who move home from one country to another usually feel frustrated when they discover that this particular standard is national rather than international, and they have to change all the plugs on their appliances. Standardisation is a powerful bridge between the supply side and the demand side of economies. It reduces uncertainty for both suppliers and buyers. This reduction occurs whether the standards are achieved through official bodies (consensual bargaining e.g. telecommunications standards) or through market forces (buyers voting with their money e.g. Betamax versus VHS for video recording).

What are the advantages of internationally agreed and ratified standards (*de jure* standards) from organisations such as ISO compared with proprietary or industry agreed standards (*de facto* standards)? There are two specific advantages and a number of more general ones. The first specific advantage is that the standard will be reviewed and modified according to a well defined time schedule, basically at four yearly intervals, and in between the standard will be stable. In contrast industry standards can be changed at the whim of the owner and, as software developers know, this can occur at short notice playing havoc with delivery schedules for derivative products. Second, official standards are designed with conformance testing in mind (many now have conformance testing schedules attached) and so buyers of derivative products have a guarantee of conformance. Such testing is rarely possible with de facto standards as the complete definition required for designing conformance tests is either non-existent or not available. The customer has thus a significantly higher risk to take when expecting to interoperate with products based on such standards. The general advantages lie in the openness of the official standardisation process. The meetings are essentially open to any companies or organisations that wish to participate and can afford to do so. The decision on whether to ratify the proposed standard is decided according to democratic voting principles on a country by country basis. There are practical problems with this process, not least the time taken to produce standards, but it is a more open process than that where one or more companies try to impose their views on the rest of the community.

ENCODING STANDARDS

Basics

Standards for information coding start at the lowest level with text characters and progress steadily to more semantically rich and complex objects. At a certain level they change from being coding of primary information such as text, pictures or documents to being coding of specific items of information such as dates, company names or countries. In principle, OII should embrace the higher level objects as well as the lower. At present we have limited the scope (at least as far as CEC activity is involved) to the lower generic coding for two main reasons. The first is one of manageability, if OII were open ended then we would have difficulty focusing on specific actions that can be completed in the short term. The second is that the higher level coding tends to be the responsibility of specific interest groups and indeed represent the coding needs of such groups which are in general independent of each other. These groups have their own organisations generally to control and disseminate information about their activities.

The scope of OII is thus focused on generic coding systems, and specifically those involved in coding what may loosely be described as documents for electronic interchange. There are three main types of information involved in interchanging documents, although all three need not be present. First, there is the *content*, the information *sui generis*. Second, there is the *structure* of the document which is particularly important in the case of interactive multimedia documents, where the content may be "read" in a variety of orders. Third, there is the *presentation information*, such as fonts for text, size for pictures, loudness for sound, scripts to control user interaction.

An important point to be made concerns the separation of information from processing. In the past the primary aim for the use of computers has been the processing of information. In the new world where telecommunications, computing and communication become merged, a primary aim is the use of computers for the communication and presentation of information. This re-orientation requires a much clearer separation between the storage and interchange of information and the processing of information than in the past. Information will be used in a variety of applications chosen by the user rather than the producer of the information. At the presentation level this distinction can become blurred due to the necessity to provide functional information as part of the presentation control. The best example of this blurring is probably the Postscript description of a document which is to a large extent a program to render the document. The information itself is embedded in the program and is very difficult to separate from the processing instructions. The two forms of information coding corresponding to separation and non-separation are generally called processable documents and final form documents. Final form documents are intended for presentation on some specific class of output devices such as laser printers or television sets. Processable documents contain information intended for further computer processing that may include editing and preparation for presentation. Due to the development of new and better presentation devices and the consequent destruction of obsolescent machines, the processable form of documents is a better form for archiving of information. On the other hand, at present the lack of market penetration of standards for processable documents means that in many cases interchange can only be achieved using final form documents, either in a programmed format such as Postcript or as bit-mapped images.

There are problems with the existing coding standards that have arisen principally due to the development of new ideas of standardisation and applications, coupled with the inherently slow rate at which agreement on standards is achieved. A common problem is a conflict between the requirement for generic standards of wide applicability that can aid interchange between different application domains, and a requirement for specific standardisation in a well defined application domain. This type of problem is demonstrated by the development of ODA and SGML (see below) to describe text document structures. ODA sets out a specific document architecture (a parametrised description of the elements making up a document, a document semantics) which is intended to cover the majority of documents in common use in business. Specific documents can be defined by the parameters, and interchanged between conformant systems with a minimum of intervention by the users *in the ODA application environment.* SGML, in contrast, is a language which can be used to define document architectures, but which does not include any specific document description. The document descriptions and associated semantics have to be developed and applied by the users in their specific application domain. The advantage of the ODA approach is that, within its specified domain, applications are developed with a well defined semantics, whereas with SGML, applications must obtain their semantics from elsewhere. The advantage of the SGML approach is that new developments in document architecture (for example adding sound annotations) do not require changes or additions to the standard, they are written into the applications by adding new semantics using the same underlying syntax. A simple analogy may be drawn with building. The buyer of a house is confronted with a choice between a pile of bricks, timbers, window frames, etc. from which he can build any style of house he desires, and several designs of pre-constructed house on a housing estate. If one of the available houses suits his requirements it is easier to buy a pre-designed house, if not he must build the house from scratch with the aid of an architect and a builder. In either case, once the house is built or bought it can be decorated and furnished according to the wishes of the owner. SGML corresponds to the pile of bricks and ODA to the housing estate approach.

This type of problem is not confined to the comparison of ODA and SGML but occurs throughout the standardisation arena. EDI standardises documents that are so specific in structure (for example invoices) that the user has little to do except fill in fields from a set of pre-set options. The aim of EDI is to so limit the semantics that the interchange of information can be carried out between computer applications without human intervention. In a sense these different standards represent a hierarchy of semantic levels that could be approached in a unified way (as indeed, *within EDI*, different business communities tackle the questions of element coding such as country codes, currency codes, addresses and dates). In practise the standards are developed in parallel by different user communities with objectives that are not sufficiently co-ordinated to ensure that the results fit together in an optimal way.

Content standards

Content standards are the standards for the lowest level of information that is to be interchanged, the level at which the data type is homogeneous. Thus the main content types are characters, images, drawings, sound and video. These content types have been the subject of extensive standardisation activity. In addition to the content types that make up documents there are a class of content types that are not consistently standardised, those that are embedded in other standards for information processing. For example, with computer languages and database management systems there are such data types as integers, floating point numbers, character strings etc. There is an effort at present to standardise these embedded content types plus newer relations such as spreadsheet information and business graphics, under the generic title "data in documents".

Characters

Character coding standards are probably the best known examples of encoding. ISO 646, ASCII in its US (ANSI) form, is a 7-bit code for the English character set. However a recent study by the European Workshop on Open Systems (EWOS) listed 62 standards for character coding. There are two main classes of character codes, 7-bit and 8-bit codes. The use of 7-bit codes is related to telecommunications restrictions, 8-bit codes make more sense for the computer community. The 7-bit code group allows in its basic representation 95 graphic characters and 33 control codes. However these can be extended through code extension techniques defined in ISO 2022 (7,8-bit code extension techniques), ISO 6429 (control functions for 7-bit and 8-bit coded character sets) and ISO 2375 (procedure for the registration of control sequences). Variations of ISO 646 supporting character sets for European languages other than English have been published over the years and work in local contexts. However as soon as one moves outside of that local environment chaos reigns and documents become unreadable. 8-bit codes, the ISO 8859 family, allow 224 characters but are generally set up for 191 characters plus 33 control characters. 8859-1 or Latin Alphabet No. 1 covers the normal requirements of most languages of Western Europe, North and South America, Australia and New Zealand. Recently a new 32-bit standard, ISO 10646, has been approved that provides sufficient space for all languages. The progressive introduction of the new standard will be a major factor in information processing developments over the next ten years. The new standard is intended to be as compatible as possible with the earlier standards and can be used in 8,16 and 32-bit formats.

Graphics

Static image information falls into two main classes on computers, vector and bit mapped. Bit mapped files describe the colour (or grey scale value) at each addressable point of an image - pixel level. Vector files contain a mathematical description of the graphic objects used to create an image, essentially the parameters required to drive the computer programme that will draw the image. This area of standardisation is particularly poor at present due to a large number of competing proprietary formats. The main vector standard is Computer Graphics Metafile (CGM - ISO 8632). A graphics metafile is a mechanism for the capture, storage and transfer of graphical information as a graphic database. In the bit mapped area one of the main standards is Tagged Image File Format (TIFF) jointly controlled by Aldus Corporation and Microsoft. There are a number of problems with TIFF that are symptomatic of this type of de facto standard. For example, the standard defines a wide range of options, and few programmes claiming to support TIFF support all of them. As a result TIFF files often fail to transfer successfully from one application to another. Another problem is that the current version of TIFF is 6.0 so that there are many applications supporting only earlier versions and unable to accept files created with later versions. On the other hand the definition of TIFF is publicly available, which is often not the case with proprietary formats. At present a new international standard for bit mapped images is being prepared, Image Processing and Interchange: Image Interchange Facility (IPI-IIF, ISO DIS 12087). The new standard is similar to TIFF but extends its scope in many ways such as allowing more than two dimensions (for time varying and multi-channel images), and defining image manipulation functions. A new standard has also been proposed for vector graphics, Presentation Environment for Multimedia Objects (PREMO). The new standard would cover applications developed since the original work on CGM, notably in the areas of animation, simultaneous use of multiple media, user interfaces, visualisation and realistic rendering. Whereas IPI-IIF is near completion and products could appear in the next 2-3 years, PREMO is unlikely to become stable in that time and products are unlikely to appear before the end of the decade.

Compression

In addition to base standards for images there are compression standards. The large size of raw image files create problems for transfer both over networks and on fixed media. At the same time it is possible to achieve significant reduction in size using mathematical compression techniques. However, these have to be standardised to achieve interoperability. During the late 1980s a great deal of work was carried out on compression standards resulting in the recent introduction of new standards with wide acceptance. There are three different standardisation groups, Joint Bi-level Image Coding Expert Group (JBIG - ISO CD 11544) dealing with images with a limited number of grey scales or colours, Joint Photographic Expert Group (JPEG - ISO 10918) dealing with grey scale and colour images, and Moving Picture Expert Group (MPEG - ISO 11172) dealing with moving images. The JPEG standard is completed and is becoming the main standard for compressing photographic images. MPEG has more than one standard, depending on the bandwidth of the target system. MPEG - 1 is completed and is intended primarily for video on CD-ROM, MPEG - 2 is nearing completion and is intended for use in broadcast digital television applications, MPEG - 3 is beginning and is intended for low bandwidth applications such as video conferencing. The current standard for video telephony/conferencing compression is H.261 of CCITT (Group XV).

Other content standards

Sound standards have been developed primarily by CCITT for use over telephone lines. The differences between them are related to distinctions between speech and other types of sounds (eg music) There are a number of standards covering different frequency ranges, different sampling rates and different compression methods. The MPEG compression standard for moving images also includes compression for the associated sound track.

In addition to standards for base content data types, there are numerous standards for high level data with defined semantics that are expressed as record structures. For example, business record standards for such items as invoices and bills of lading are covered by Electronic Data Interchange standards (EDI - ISO 9735), the exchange of product description data widely used in engineering is covered by Initial Graphic Exchange Specification (IGES) and work is underway for a new standard in this area (STEP - Standard for the exchange of product data), standards exist for bibliographic and statistical data of many types and work has begun to standardise geographic information at an international level.

Document standards

Interchanging whole documents or sets of documents raises a number of new problems. In addition to the text itself, and any embedded images, there is information about structure (headings, paragraphs etc.) and about presentation (page layout, fonts, etc.). This information is stored in word processing or desk top publishing files in proprietary formats which initially were unavailable to developers, and then when they were released (by some vendors) were changed at such frequent intervals that keeping converters up-to-date was almost impossible. Two groups of standards developers have worked on solutions to this problem from different directions. One group, Open Document Architecture (ODA - ISO 8613), set out to create a standard for interchange that would enable transparent interchange of documents between word processing and similar applications. The solution expected the vendors of the applications to provide an ODA output format as part of the package. The other group, coming from a publishing background, set out to standardise mark-up (a traditional method of embedding process information into publication information) through a tag syntax that could be used for many applications but did not provide any semantics (Standard generalised mark-up language - SGML - ISO 8879). In theory ODA is better for users who wish to interchange documents because they should not have to "see" ODA working. In practise the complexity of ODA has limited the number of vendors prepared to implement it and it has proved easier in many cases to use SGML with locally defined semantics. SGML is essentially a method for adding processable information into a document in a standard way. This information may be related to presentation, but need not be and indeed in many applications is not. The comparative success of SGML to date is certainly due to this factor rather than any comparison with ODA in terms of presentation interchange facilities. SGML tagging provides a convenient way to transfer information from one form to another through an intermediate form that is relatively application independent, for example in the conversion of EDI messages to database input records. Tagging also enables information to be interchanged without specifying the final presentation form of structure features, for example headings might be presented in a printed form as separate items with visual emphasis (bold or italics) but in a screen based presentation as a menu list.

The recent growth in interactive presentation of information has led to new forms of document structure. The principal change is in the development of non-linear access paths in documents, commonly called hypertext. Hypertext creates links between different parts of a document that can be used in menu choice situations to enable the reader to jump to (or selectively display) new sections of a document. In order to interchange documents containing this type of structure new standards have been developed. The main standard at present is HyTime (ISO 10744) which is an extension (application) of SGML. Another new aspect of document structure is to include information other than text and static pictures, mainly video and sound - multimedia. Extensions to ODA are envisaged for both these types of applications. In addition the Multimedia/hypermedia Information Coding Expert Group (MHEG - ISO DIS 13522) is developing a standard for the interchange of multimedia objects in real time interactive systems and in final form interchanges. Work has also started on the standardisation of scripts to drive interactive multimedia document applications, Standard Multimedia Scripting Language (SMSL).

If SGML provides a means of describing document structure independently of presentation, then standards are needed to convey presentation information where that is also required. This information may be interchanged as additional information or the final form presentation object may be interchanged directly. Information to connect tags in the document with presentation features is standardised in Document Style Semantics and Specification Language (DSSSL - ISO DIS 10179). DSSSL is a draft standard that has been through many revisions and is still not clearly stable. At the presentation level, documents are often exchanged using the de facto standard Postscript of Adobe. A new international standard, Standard Page Description Language (SPDL - ISO 10180), based on Postscript has recently been completed. It is not clear at present whether this standard will replace Postscript in products. Other aspects of presentation such as fonts have been standardised (Font information interchange - ISO 9541).

APPLICATIONS

It would be foolish to suggest that data and documentation interchange standards are easily applicable at present. One of the principal problems is a lack of awareness, amongst developers and users, of the standards available and how to use them. As a result many of the most popular application packages do not support the existing interchange standards. Often the vendors or users claim that the proprietary formats they use provide specific facilities that are not in the standard, but this is rarely true. For example, a recent study of graphic standards carried out by PIRA for the OII activity found that categorising standards by application domain was unsatisfactory, many of the standards cross application boundaries and that the new IPI-IIF standard could cover many application areas.

The major successes in the application of information interchange standards are in areas where the volume of information and the size of applications are sufficient to justify major investment. These tend to be projects of government and large corporations. The success of SGML in document interchange is principally in large corporations which have technical documentation systems, for training and maintenance manuals, that can justify automation and distributing reader stations to clients. In the EEC, groups specifying standards for public procurement are beginning to include document interchange standards. The next edition of the European Handbook for Open Systems (EPHOS 2) will include guidelines for ODA and SGML procurement. The US Department of Defence CALS initiative helped boost the number of systems supporting SGML and this enabled companies to buy solutions to their own electronic publishing problems. Examples can be found in papers given at SGML and technical documentation conferences organised by the GCA, companies such as Ericsson, Shell Exploration and Bell Helicopters. One of the lessons that has been learnt, is that introducing standards for document interchange is a major organisational task, at the same level of complexity as introducing database management systems was in the 1970s [5,6]. Applications have to be developed to suit the individual company, and whilst they can be built around standard packages to support document authoring, document storage and document dissemination, the system integration costs are high and extensive organisational planning is required. Publishers are also moving to use SGML, mainly to handle the creation of alternative products from an electronic information base. Cheap viewers, for electronic books distributed with SGML mark-up, that can drive hypertext access and handle the presentation aspects of different platforms are becoming available. ODA is seen primarily as an application in OSI environments, so that its success is tied to that of OSI itself

An example of a major international research project is DIDOS, R 2037 [7] from the CEC's RACE programme for advanced telecommunications applications. DIDOS will specify and realise distributed applications for the production of technical documentation based on a service centre concept. The main objective of the service centre concept is to offer TD-producers and service providers an efficient way to distribute the production of technical documentation across a network. The service centre concept also enables customers to gain access to a network of distributed services from a single local point of access. The TD-producer sends documents to the service centre electronically, the user orders copies from the service centre and the service provider organises the whole operation. There are three application pilot projects in DIDOS. Application pilot 1 is concerned with the composition of technical documentation. A document application design service (DAD) will be demonstrated that supports the design, configuration and implementation of composing tools and services. The application includes a DAD workbench of tools and explores the practicality of using "software rental" as a means of controlling the cost of highly variable use of high-cost document layout and management tools. Standards used include SGML and SPDL. Application pilot 2, is aimed at a "Just-in-time" production environment for technical documentation. The project is concerned with closing the gap between Computer Integrated Manufacture and the need for finished documentation in time with the product, the add-on-value of timely documentation, supporting the sale of products in many variants with documentation to match both the specific product specification and the user language requirement. The project also uses printing on demand with remote printing facilities. Standards used include Postscript, TIFF, EDI, ISDN and SQL. Application pilot 3, will implement and demonstrate a distributed multimedia service centre. The services provided include management, design, capturing, converting, CD-ROM production, hypermedia production, distribution and on-line databases. Standards used include SGML, HyTime, TCP/IP, and FTP. Overall the project partners are Detecon, Fogra, Tekom, and Bertelsman from Germany, PIRA, and Crosfield from the UK, DTI and Grundfos from Denmark, DEC from France, and Intracom and Ntua from Greece.

CONCLUSIONS

In this paper I have tried to give an overview of the importance of standards for document interchange and an idea of where we are at present. There is a growing demand for documents in an electronic form to be interchanged between different platforms and software environments. Standards exist to enable most information to be interchanged in an application neutral way. The main barrier to such interchange is lack of awareness of the methods available and a lack of communication between users and vendors as to the priority of improving interchange. Users have to make their priorities clear to vendors if they are to be offered the facilities they require in the application packages they want to use.

For further information:

Copies of standards and EWOS reports are normally only available through national standards bodies. Character standards prior to the new 10646 standard are well described in EWOS report PT N001: "Usage of coded character sets and repertoire in EWOS". Information on graphics standards is available in the PIRA report "Open Information Interchange study on image / graphics standards", available (as is further information on OII and IMPACT) from IMPACT Central Office DG XIII / E, L-2920 Luxembourg, fax: +352 4301 32847. The results of a recent workshop on hypermedia standards sponsored by the CEC OII initiative was published in a special August 1993 edition of the journal "Information Services and Use" (IOS Press): Proceedings of the 1st Workshop on Hypermedia and Hypertext Standards. Copies can be ordered from: Johan van Halm Information Consultancy, P.O.Box 688, NL-3800 AR Amersfoort, Netherlands Fax: +31 33650945, the price is 100. DFl (incl. postage, official programme, list of participants and other documents distributed during the workshop). A book has recently been published on JPEG "JPEG still image data compression standard", Mitchell and Pennebaker, Van Nostrand Reinhold, ISBN 0-442-01272-1. An introduction to SGML in the form of a hypertext driven from an SGML marked up file with its own viewer is "SGML tutorial", Eric van Herwijnen, Electronic Book Technologies, One Richmond Square, Providence RI 02906 USA, fax: +1 401 421 9551. A study of the relation between ODA and SGML is published by EWOS as EWOS PT N011: "SGML / ODA convergence". A study of multimedia "A survey of distributed multimedia research, standards and products", C.Adie (ed.) was carried out by Edinburgh University Computing Centre for RARE (Reseaux Associes pour la Recherche Europenne) early in 1993. This report, which is comprehensive and full of interesting information, is available by anonymous ftp on Internet from ftp.edinburgh.ac.uk:pub/mmsurvey. A number of papers on multimedia standards (JBIG, JPEG, MPEG, MHEG) appear in the proceedings of a workshop held by AFNOR in Paris in January 1992, "Multimedia et normalisation", AFNOR, 1992.

References:

1. Stephenson, G.A., "Open information interchange", Inf. Services & Use, *12*, 1992, pp 235-246.

2. Johnson, T. and Chappell, C., "The computer integrated company: market driver for the 1990's", OVUM, 1990.

3. Moore, N. and Steele, J. "Information intensive Britain", Policy Studies Institute, 1991,

4. Ramfors, G. "Technical information as an integrated part of a product or a function", *in* Documentation Europe '92, Graphic Communications Association, May 1992.

5. Price, L., "Managing the organizational change caused by SGML", *in* International Markup '92, Graphic Communications Association, May 1992.

6. Mulder, K., "Is SGML bad news to your editorial staff?", *in* International Markup '92, Graphic Communications Association, May 1992.

7. Cave, F., "Just-in-time printing of technical documents in a distributed environment", *in* Documentation Europe '93, Graphic Communications Association, May 1993.

The Virtual Library: Coming of Age

Judy F. Hunter
Gladys A. Cotter
National Aeronautics and Space Administration
Scientific and Technical Information Program
Washington, D.C. 20546
U.S.A.

SUMMARY

With the high speed networking capabilities, multiple media options, and massive amounts of information that exist in electronic format today, the concept of a "virtual" library or "library without walls" is becoming viable. In the virtual library environment, the information processed goes beyond the traditional definition of documents to include the results of scientific and technical research and development (reports, software, data) recorded in any format or media: electronic, audio, video, or scanned images. Network access to information must include tools to help locate information sources and navigate the networks to connect to the sources, as well as methods to extract the relevant information. Graphical User Interfaces (GUIs) that are intuitive and navigational tools such as Intelligent Gateway Processors (IGP) will provide users with seamless and transparent use of high speed networks to access, organize, and manage information. Traditional libraries will become points of electronic access to information on multiple medias. The emphasis will be towards unique collections of information at each library rather than entire collections at every library. It is no longer a question of whether there is enough information available; it is more a question of how to manage the vast volumes of information. The future equation will involve being able to organize knowledge, manage information, and provide access at the point of origin.

1. INTRODUCTION

Largely due to advances in technology, organizations are transitioning away from mainframe-supported computing environments towards distributed computing environments. The trend is to decentralize execution, but centralize the management for the establishment of policies and budgets. Due to the globalization of industry and a competitive world market, organizations demand information that is generated from their internal processes as well as external information, such as reports about research conducted in other countries.

The concept of a "virtual library" or a "library without walls" is meant to convey the idea that information in any format is an information resource and it should be available to the enduser from the desktop as if it were located in the local library. The physical, geographic location of the information and the techniques used to access it are not relevant from the enduser perspective. The enduser needs information and tools to exploit it. Everything else should be transparent. The idea is to be functionally and geographically distributed, but logically centralized.

The virtual library concept is possible given the high speed networking capabilities, multiple media options, and massive amounts of information that exist in electronic format today. Distributed networks and databases allow for distributed architectures. In the virtual library environment, the information processed goes beyond the traditional definition of documents to include those recorded in any format or media: electronic, audio, video, or scanned images. Network access to information must include tools to help locate information sources and navigate the networks to connect to the sources, as well as methods to extract the relevant information. Graphical User Interfaces (GUIs) that are intuitive and navigational tools such as Intelligent Gateway Processors (IGP) will provide endusers with seamless and transparent use of high speed networks to access, organize, and manage information.

This paper focuses on virtual library concepts and technology. It uses the NASA Scientific and Technical Information (STI) Program for specific examples. NASA is one of the major research and development agencies of the U.S. Federal government. The NASA STI Program was established as part of the Space Act of 1958. Its mission is to identify world-wide sources of scientific, technical, engineering, and related information, develop required policy statements, facilitate authorized access, and manage delivery of the information to NASA and its customer base. The NASA STI customers include NASA employees, NASA contractors, other U.S. government agencies, other U.S. government agencies' contractors, the U.S. education system (K-12 and universities), international partners, and the general public.

The widely accepted definition of STI within the Federal government is: "(1) basic and applied research that results from the efforts of scientists and engineers (including new theory and information obtained from experimentation, observation, instrumentation or computation in the form of text, numeric data or images) and which may be further transformed, described, evaluated, synthesized and recorded in print, digital, magnetic or other media to enhance its communication and its usefulness and value to a wide spectrum of endusers and uses, and (2) information that bears on business and industry generally, such as economic information, market information and related information, if the agency determines such information would be of value to consumers of the information described in the preceding subparagraph."(1) The underlying importance of this function is to prevent the duplication of previous efforts for time, effort, and expenditure of resources.

With the current U.S. policy's emphasis on global competitiveness and the transfer of technology from the Federal to the private sector, NASA, like the other Federal agencies, must be able to identify sources of STI and make the information available to those who need it in a usable and timely manner. The Federal government is not unique in this requirement. Private industry is struggling to maintain its competitive edge within the world market. Corporate Executive Officers (CEOs) in private corporations consider information critical to the success of their businesses. Successful businesses will set up organizations that optimize

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

global resources. The successful management of information as a resource will be a critical success factor for businesses in the 21st century.(2) Success will be based on information literacy; that is, the ability to find and use information to construct a competitive advantage.

2. FUTURE VISION

The critical component in the "virtual library" concept is the part that creates the illusion for the enduser of having a "one stop shopping" capability. In order to do this, technology and information specialists are needed. Ultimately, the goal is to create a scenario described below:

"A NASA scientist, Dr. Bennett, sits at his desk where he has docked his personal information server at its docking station. The docking station provides connections to the Center LAN [Local Area Network] and the NREN [National Research and Education Network] with its wide variety of information resources. A chime sounds and a synthesized voice emanates from the docking station reminding Dr. Bennett that he must prepare a report today for his section head on his planned areas of research in order to begin the process of identifying funds for his efforts.

Although Dr. Bennett finishes most of the report based on extensions of work he has done in the past, he wants to include something new as well. He has been thinking about identifying a set of volcanic gases and using their compositions to predict the effects of volcanic gases on the Earth's protective ozone layer. Dr. Bennett selects the natural language mode from among the various interface modes available (icon, voice, menu) to query the virtual library of scientific and technical information available to him.

Dr. Bennett doesn't specify whether he wants a full text or a keyword search. The system knows, from checking its knowledge base about this use, that he prefers keyword searching and [that he] typically requests peer information. Dr. Bennett types in his natural language query: ***What research has been done or is in progress on the effect of volcanic gases on the Earth's ozone layer?*** *The query is automatically expanded to include all relevant keywords. When the query is entered, the peer locator identifies all researchers working in the area of the query and adds an icon to Dr. Bennett's screen to indicate that peer information is available.*

Dr. Bennett, as a NASA researcher, isn't surprised when the first result of his search is a surrogate for a NASA STI program "live" multimedia document. Dr. Bennett reads the title of the first document and views the thumbnail images of Merlin, an eight-legged robot, and Mount Rebus, a remote volcano. He clicks a button to indicate that he wants to hear rather than read the abstract, which clearly states that the document presents Merlin's retrieval of volcanic gases, and measurements of gas composition, and ties the information to potential effects of the gases on the ozone layer. He clicks again to bring up the live document and begins by watching a narrated video clip of Merlin's descent into the mouth of Mount Rebus.

He then reviews the composition of several volcanic gas samples that Merlin collected and requests that these samples be contrasted and compared with known ozone depleters. He next requests that the provided software tool both predict and graph the effect on the ozone layer over time for several gases most similar to known depleters. He also specifies several release quantities for use in the model. The software begins its task when an ALERT icon appears on the bottom of Dr. Bennett's high resolution display. He clicks on the ALERT and is pleased to receive news that a 20 minute live video feed from Merlin will begin in 30 seconds. This really is research in progress!

Live, Merlin quickly identifies a previously unidentified gas from one of the volcano's vents. When Merlin analyzes the composition, Dr. Bennett immediately recognizes the gas as related to a severe ozone depleter that had been synthesized and used in petroleum refining several years ago. He adds a voice annotation to that effect to the document. He finishes viewing the video and changes the time and quantity variables used in the ozone layer depletion predictive modeling. He clicks on FORMULATE and views the computations used to create the model.

He then reviews the surrogates for the other documents that resulted from his search. There aren't many because this is a relatively new area of research; clicking on the comprehensive bibliography icon doesn't bring up many related documents either. He changes a setting on the bibliography icon to switch from English only to any language and clicks to view the expanded bibliography. The English title to a French document looks appropriate, so he responds yes when asked if he wants a rough translation. Dr. Bennett downloads the French document, the Merlin document, and the peer review information to his personal information server storing it on his magneto-optical drive. The system updates Dr. Bennett's entry in the knowledge base and adds the new search topic to his areas of interest. Dr. Bennett then uses the peer information to establish contact with Dr. Smith, principal Merlin researcher. The conversation indicates that a collaborative effort may be in both researchers' futures...

Dr. Smith has created his own live document, including digital video clips, using standard desktop tools of the day. Software on the Center LAN automatically routes a copy of the document to the STI Program where it is input to the virtual library catalog and document retrieval system. A copy is also forwarded to and retained at Dr. Smith's Center. The STI Program determines the interchange standards and maintains the virtual library union catalog and comprehensive set of documents from all Centers. A researcher may use the Program's virtual library and document retrieval system to identify a document and request that the information be transmitted electronically or mailed. The researcher may choose to have a print image sent to his or her site for production on a local printer."(3)

3. TODAY'S REALITY

Although this scenario is not in place in its entirety today, it is not as far away from implementation as it may appear initially. Many large, respectable American corporations are incorporating similar goals in their Information Resources Management (IRM) strategic plans. Boeing Computer Services, for example, states "Computing suppliers are gradually putting together an infrastructure which will allow electronic information to be readily, conveniently, and easily accessed and sent to others. The infrastructure will be the next century's information equivalent to today's worldwide voice telephone system."(4) The Apple Corporation has a video that proposes a enduser scenario similar to the one described in the

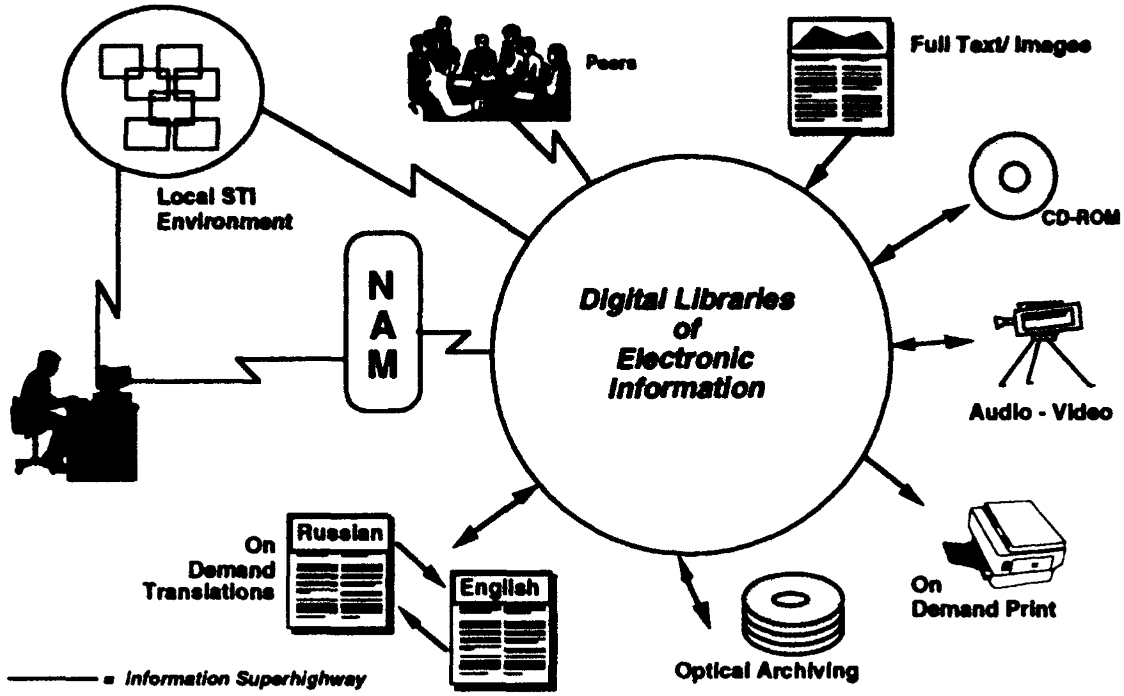


Figure 1. The Virtual Library Concept

Section 2.(5) This video was created in 1987 and Apple is serious about working towards implementation. In the current version System 7 of the Macintosh operating system, Apple has included a "Play Movie" option which is a precursor to being able to create and render multimedia documents. Federal agencies have been designing and prototyping "intelligent gateway" systems to provide their enduser communities with "one stop shopping" since the mid-1980s.

The technology is available today, though in various stages of development, to establish the environment shown in Figure 1. Most endusers are connected to a local area network (LAN) within their organization. The functions on the LAN include a word processing package, a mechanism to send and receive electronic mail, a spreadsheet package, and maybe even a database management system. Also situated in the local environment are sources of information that are generated from within the organization, such as data on a mainframe computer or on local file servers that are accessible and may be shared by all or by designated segments of the organization. For information that is required from sources outside of the organization, the LAN and/or the enduser workstation may be linked via telecommunications to other LANs or Wide Area Networks (WANs), such as the Internet, or via direct dial-up lines to specific hosts. For the purposes of this paper, all of the information resources that are accessible using any access methodologies (LAN/WAN technology, direct dial-up methods) are identified as libraries of information.
Historically when a enduser (student, researcher, CEO) needed information from outside, a trip to a library was essential. A library was a physical building. In this scenario, it is assumed that much of the information needed today is either in an electronic format or there are electronic records, such as those contained in a bibliographic database, that point to the sources available in a paper format. In any case, most document ordering facilities are electronic. This then is the virtual library.

The heart of the virtual library is the collections of electronic information. In the virtual library schema, such collections are a part of the information infrastructure. That is, they are connected to the communications infrastructure in some way and are accessible electronically. These collections of information or "libraries" include people resources, which are being recognized more and more as a critical information resource. In the research environment, for example, the scientists communicate on a regular basis with a small, select group of experts in the same discipline. To be accessible through the network, the people require an electronic mail address or a reference within the network someplace to their telephone number or postal address. The idea is to facilitate communications. If this means providing a telephone number—access has been facilitated successfully. The better application is to be able to send the "resource" electronic mail or even to transfer a full file to that person for review.

The digital libraries include stores of complex documents that contain fulltext and images; information located on CD-ROMs, audio and video, and other optical media. Especially in the research environment where the scientists are using sophisticated modeling techniques on super computers to perform experiments and conduct research, they expect to be able to get data in various formats on various media in foreign languages.

The goal is for the enduser to have immediate access to relevant information in any format from the desktop. It is not enough to give the enduser all the information that is available on his topic. The requirement is to provide *only* the specific information needed by the user. In the past, the enduser was flooded with all the information ever created about a particular subject. The enduser didn't have the time or the patience to sift through massive volumes of information to identify those pieces that were relevant. The enduser needed to be in control of the information as opposed to the information controlling the enduser.

Key to the success of the virtual library concept are tools to identify sources of information, navigate the communications infrastructure to get access to the sources, and analyze the data for specific requirements. In Figure 1, the NAM (NASA Access Mechanism) box represents a graphical enduser interface (GUI) and an intelligent gateway processor (IGP); which is the first point of access for the endusers to the information resources. The GUI must be intuitively easy to learn, assist the endusers' identification of relevant sources of information, facilitate access to those resources, and provide tools for analysis of the information.

4. BENEFITS

The major benefits that can be realized from a "virtual library" include the ability for the endusers to use information as a tool in their decision making. Until now, information has taken on a life of its own.

The information can be collected and maintained by the owner. Individual collections of information will be specialized rather than everyone expending resources to collect everything. Other information will be accessed at the point of origin using high speed networks and navigational tools.

Information is available immediately in a format that is usable to many endusers. In traditional libraries once a book is borrowed, it is unavailable for others until it is returned. In the virtual library, one "book" can be read by many at the same time.

The concept of "one stop shopping" is a major benefit to the enduser. Historically, the enduser has had to know about a source of information and then be able to figure out how to gain access to the information. Once the information was obtained, the enduser then had to put the data in a format that was usable to him for his application. If the enduser gained access to a database containing information, he then had to learn the query language in order to retrieve data from the system. The learning curve was not insignificant. In the client server architectures and GUIs that are being developed today, it is possible to create an environment for the enduser using "point [a mouse] and click [on the icon]" methodologies to simplify functions and mask the complexities from the enduser. Figure 2 depicts the main menu of the NAM prototype system developed in the last two years at NASA.

The NAM functions include a source locator that searches a database of the available sources to select those sources that cover the subject of interest to the enduser. If the enduser knows the database he wants to query, he can choose that host as one of the options available within the **Data Sources** button. On the other hand, the enduser may not know which database contains the information he needs. In this instance, the enduser may enter a particular area of interest, such as *aerospace sciences*, and the system will recommend the database that contains the most information about aerospace sciences. It is within the **Data Sources** function that the connections to the host computers occurs—automatically and transparently. The system is designed to "know" which communications paths are available to reach a particular host and includes a list of priorities. That is, if the Internet connection is not working, the system will try an X.25 connection or a direct dial-up connection.

The **EMAIL** function gives the enduser access to an electronic messaging system that he can use to send electronic mail to other users around the world. This function can also be used to mail any information that is downloaded from a host computer to himself or a peer.

The **PEER LOCATOR** function facilitates communication between peers (colleagues). The function is provided using a database that contains the name, address, telephone number, and electronic mail address of people. For the NAM prototype, the digital telephone books were collected from 13 NASA Centers and loaded into an INGRES database management system. For people external to NASA, an Internet utility called **FINGER** was added that will search all hosts on the Internet (the host must be running Finger) to see if the person in question has an account. Late in the prototype, NASA added an implementation of X.500, which is a locator standard that is part of the Open Systems Interconnection (OSI) suite most commonly referred to as the "White Pages." X.500 is designed to serve as the electronic telephone book for the world. NAM will automatically insert the electronic mail address into the electronic messaging system to facilitate sending mail.

In the research environment, scientists rely on information they exchange with their peers informally (first and foremost), on information that has been published in journals and has met a peer review requirements for publication (secondly), and finally on information they may find someplace that they can

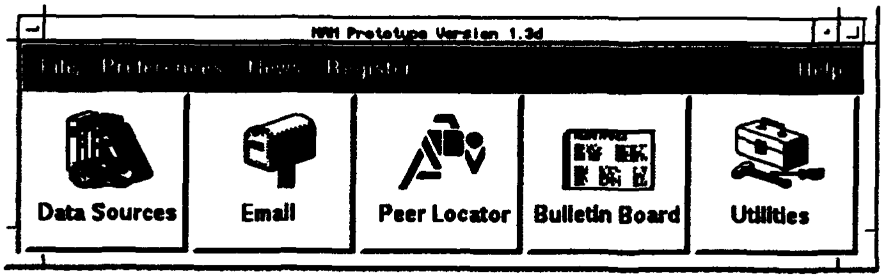


Figure 2. Prototype NAM Main Menu

validate themselves if it seems relevant. The **Bulletin Board** button includes access to volumes of information located on the Internet on public and private file servers that fall into the third category. In some cases, the enduser wants to get a general idea about any information that is available in a certain area. The endusers need tools to analyze the information. In most environments, some type of 3-D modeling capabilities are required. The **Utilities** button provides the enduser with tools to assist in the analysis of the information that has been retrieved.

These facilities are all provided through an iconic, form-based interface, where the user fills in a form with the keyboard or selects options with the mouse to prepare a query. The NAM handles connection to the remote system, selection of the appropriate application, and processes the query at the remote host. The NAM also presents the results of the remote session to the enduser in windows that allow the user to perform further action on the retrieved information.(6)

The NAM is one example of how it is possible with the technology available today to take advantage of the client server architecture to logically centralize geographically distributed applications and information.

5. CONCERNS

There are significant challenges associated with implementing a worldwide virtual library—critical success factors. There are cultural changes that will have to be managed across the information enduser community as well as the information development and management communities. The introduction and proliferation of the information age will shake some traditional processes and mindsets. People resist change.

The development of and adherence to standards will be one critical success factor. International communication and distributed computing standards are essential to the transfer of information among organizations or countries. Standards development is slow; therefore, cooperation among vendors is crucial to ensure interoperability between multi-vendor platforms. Though some flexibility is required to be able to develop outside of standards, the time of "Rambo engineering" where a system is developed quickly without any plan for the future is past.

The networking infrastructure is another key element. In the United States, the Internet is heavily used. The Internet is the name used for the network of networks that enable computers worldwide to communicate. It was developed by the Department of Defense Advanced Research Projects Agency in the late 1960s. Though the U.S. is the largest enduser, the international connectivity is rapidly growing. "In January 1993, there are 10 million people from 102 countries connected to the 9000+ networks that comprise the Internet. The number of people is continuing to double annually."(7) The Internet is the foundation for the National Research and Education Network (NREN) initiative described as the backbone of the National Information Infrastructure being promoted by U.S. Vice President Gore.(8) It is this backbone that is the basis for the "glassing of America," which involves laying fiber optic cable to handle high-speed, high-volume transfer of information across America.

Finally, a major challenge is that of intellectual property rights. How can you control copyright issues on electronic documents? How will authors and publishers receive payment for their work? How will the role of the publishers change in the new environment? These are but a few of the questions being asked, especially by the publishing community, in forums across the world today. Questions regarding intellectual property rights in regards to electronic document transfer and multimedia are just beginning to be addressed. "Every time a new technology comes along, there's a period of 10 to 15 years with litigation. It happened with movies. There was a whole series of cases when TV came along, then cable TV; now with videodiscs and CD."(9)

Another issue is how is electronic information archived? Questions have been raised about the life expectancy of digital information. Will it be impossible to retrieve data archived on optical media in twenty years due to advances in computers and operating systems? Ten years? Five years? Rumor has it that the Boeing Corporation has a couple of Cray super computers in moth balls to be sure that they will be able to retrieve the engineering data they have archived on optical media. Somehow we must be able to protect our ability to retrieve the massive volumes of information contained on multimedia.

6. CONCLUSION

In the past, data processing professionals and information retrieval specialists have been concerned with the care and feeding of the data from a "data" perspective. Today, we have moved (are moving) into an environment where information is abundant in massive volumes and the key is to identify and access relevant information. The emphasis is on the enduser. The focus is on the customer. Technology is the enabler; the customer is the driver. The key is to allow the requirements of the customers (endusers) to push the development and not to allow the technology to push the development.

The future will be the time of the "knowledge worker," that person who knows the information. The current paradigm shift is where more and more applications programmers are moving from the information system to the functional areas. Distributed systems are so popular because people have a strong interest in being in control of those things for which they are responsible.

Traditional libraries will become points of electronic access to information on multiple medias. The emphasis will be towards unique collections of information at each library rather than entire collections at every library. It is no longer a question of whether there is enough information available, it is more a question of how to manage the vast volumes of information once it is located. The future equation will involve being able to organize knowledge, manage information, and provide access at the point of origin.

7. BIBLIOGRAPHY

(1) "Part 1180—Transfer by Federal Agencies of Scientific, Technical and Engineering Information to the National Technical Information Service," Federal Register, Volume 58, Number 89, Tuesday, May 11, 1993, Proposed Rules.

(2) Kanter, Jerome, "Managing with Information", Prentice-Hall, Englewood Cliffs, NJ, 1992.

(3) Bauldock, Barbara; Ostergaard, Kristin; Kaye, Karen, "NASA Scientific and Technical Information Program Infrastructure Upgrade Plan", NASA STI Program, February 1, 1993.

(4) Boeing Computer Services, "Long-Range Technology Forecast", February 1993.

(5) Apple Computer, Inc., "Knowledge Navigator", Video, January 22, 1987.

(6) Hunter, J.; Duncan, D.; Generous, C., "NASA Access Mechanism—Graphical User Interface Information Retrieval System", NASA Technical Memorandum 108981, January 1993.

(7) Markoff, John, "Building the Electronic Superhighway", New York Times, Sunday, January 24, 1993.

(8) Clinton, William J.; Gore, Albert, Jr., "Technology for America's Economic Growth, A New Direction to Build Economic Strength".

(9) Interactive Multimedia White Paper: "Promise and Pitfalls", Knowledge Industry Publications, Inc., 1991.

PROBLEMS OF INTERNATIONAL EXCHANGE OF SCIENTIFIC AND TECHNICAL INFORMATION

FOR A RUSSIAN AEROSPACE RESEARCH INSTITUTE

VICTOR M. TYURIN
Deputy Director
Central Aerohydrodynamic Institute (TsAGI),

1, Zhukovsky street,
Zhukovsky, Moscow region,
140160, Russia

Summary

The problems of international exchange of scientific and technical information for an aerospace research institute in Russia arise from this research area being classified; publications were strictly limited. This resulted in absence of specialized data bases; there is a very little amount of information prepared properly for utilization in electronic networks. The other problem is underdevelopment of domestic networks for information exchange; there are not channels for connection with international networks. Changes in Russian policy and economics necessitate settling these problems for increasing efficiency of the organizations. There exists a plan to develop multibranch high-speed information-exchange networks, each having a coordination/control center connected with international networks.

* * *

The TsAGI is a major research center of the aircraft industry in Russia. It has been founded in 1918, and for the 75 year period has gained a unique experience in basic and applied studies and engineering for aerospace applications.

Main areas of activities of TsAGI include.

1. Fundamental research:
   - aerodynamics of all types of aircraft at speeds ranging from low subsonic to orbital speeds;
   - airframe strength (static and fatigue strength, aeroelasticity), structural materials;
   - flight dynamics, flying vehicle control systems;
   - hydrodynamics;
   - aeroacoustics;
   - industrial aerodynamics (including wind-energy plants, fans, etc.);
   - unique test facilities and data acquisition systems.
2. Exploratory research:
   - predicting the development of aerospace technology;
   - advancing the aircraft and spacecraft concepts.
3. Applied research:
   - designing specific aerospace technology prototypes;
   - scientific support of all the projects carried out at Design Bureaus, from preliminary design stages through to aircraft manufacture and flight testing.
4. State examination:
   - the institute gives resolutions on all aircraft designs, on readiness for flight tests, on service life limits;
   - the institute participates in working out all standards associated with aircraft development, testing, certification, and operation.

To implement its functions, the TsAGI has developed the unique test facilities and the pilot production division which have no close analogues world over. The institute employs some 8000 people, a half of them are scientists and engineers.

The institute is located in the town of Zhukovsky, forty kilometers from Moscow; it has a branch in Moscow and is assigned faculties at Moscow Physico-Technical Institute and Moscow Aviation Institute.

Being a leader in the aircraft industry, TsAGI coordinates acquisition, processing, and propagation of scientific-technical information in the industry; the institute has a specialized service and a small-size printing office.

The main problem of the information exchange arises from the information being classified. It was traditional that organizations of the branch were working for customers within this country only and were financed by the State.

Therefore published were mainly the theoretical and experimental results rather general and not related to specific technologies. In respect of the amount, these were "on top of an iceberg" – note that not the most valuable portion at a time instant. All the remaining work was not permitted to be published and was propagated through classified publication channels (except for software programs) and to strictly limited groups of specialists. Placed in the common area was information from neighbouring areas in science and technology. The general-purpose data and considerations were and are being included in comprehensive data bases.

Information about Russian efforts in the aerospace field was systematized poorly, and specialized data bases were not created. A few, specialized conferences published the classified collections of works that were summarizing. More systematized and more easily available was patent information.

Of course, these conditions prevented development of information bases; they were not interconnected by networks and were not designed for international use. Currently, Russian organizations have no prepared information ready for being read by foreigners.

What has changed for the last years and why does participation of such establishments as TsAGI in international information exchange become possible?

The favourable change in an international climate has significantly influenced the orientation of investigations in science and technology, in the aviation branch. Budgetary financing is being reduced and directed to investigations in civil aviation and to conversion applications of the potential of the aerospace industry. Many options prove to be owned by customers and contractors, not the State, and get commercially valuable. Information on these options must be brought to potential customers all over the world. Being completely or partly on a self-finance basis, such centers as TsAGI, owned by the State, may receive permissions from the Ministry of Defence to exchange and sale most technologies previously prevented from publication. Abstracts of such works in English may be transferred to machine-readable media and this work may be carried out in quite short terms and begun immediately.

On the other hand, many scientific results of leading specialists become not used because of reduction in the amount of orders within the CIS. These efforts are qualified highly, they are often performed using unique facilities and equipment, and may be successfully used abroad. International scientific and technical conferences conducted recently with participation of Russian scientists, certainly indicate the necessity of wide-range and intensive exchange of information. Our scientists and engineers are interested in their works becoming known all over the world, and want to have a "hot-line" access to the specialized foreign information. In addition, it is necessary to have effective electronic international network for working on the international contracts currently involving the institute. TsAGI has concluded for two last years a set of contracts with scientific centers and companies of the U.S.A. (NASA, Boeing, Lockheed, Rockwell, etc.), France (Aerospatiale, ONERA, Dassault, etc.), Germany (DASA), Great Britain (BAe), China, South Korea, India, and other countries. The number of such links grows permanently.

Due to its geographic position, TsAGI integrates with many organizations of both the defence industry and civil branches located in Zhukovsky (that is a center of aircraft industry) and, more widely, in the south-eastern region beneath Moscow. Most of these organizations are involved into conversion and diversification of their activities, therefore they would like to establish highly-specialized exchange, to promote the exchange of common scientific and technical information, and to become users of domestic and international networks.

At present, Russia and other CIS countries do not have relevant networks capable of working in batch mode and accessing foreign data channels. No organizations in the defence industry complex are subscribers of such systems.

For efficient functioning in the present economic relations the organizations must have come to a qualitatively new information support based on the modern means of communication and computers.

The analysis of the modern concepts "mass media" and "multi-media", trends in the development of information technologies in industrial countries indicates, first of all, an extremely high level of complexity of the problem under study. This, in fact, relates not so much to the traditional, human-support-oriented means (telephone, fax machines) as to the computerized information transfer communications. The modern information systems comprise the following main groups:

1) sources of information or data bases (DB),
2) information transfer means (machine-readable media, networks, etc.),
3) terminals of the users for information reception and transfer,
4) languages for users to interact with information sources,
5) subsystems for analytical processing of the information received.

The really operating systems of data exchange combine in various ways the options of the above five groups, thus offering a great number of combinations and technical solutions. This variety of versions places an individual user in a situation when he finds it difficult to make the right decision and identify proper means and techniques.

Aggravating factors for our country include not only a significant lag behind the western countries in this problem but also the steps that have been already done or are being made in a faulty attempt to get over this gap "by a jump."

To illustrate the above, one may address such an example. The computer-based "informatization" of the West started with large data bases and networks for accessing them. In the process of their development, standards have been defined for options to be used when working with remote DBs. Only thereafter the personal computers (PCs) were introduced instead of remote terminals, considerably increasing the analytical capability of the final user in what concerns processing the information obtained from the DBs.

Unfortunately, for some reasons Russia has not joined the epoch of large information centers with comprehensive means for remote access. Moreover, in recent years, due to changes in the State policy and a "personal computer boom" the electronic data processing centers reduce or cease their activities. Instead, on the basis of their own isolated personal computers, the programmers in the uncoordinated efforts start to create more and more powerful information-systems with their non-unified and non-standardized languages and non-compatible options for machine-to-machine exchange.

This is the very case the TsAGI is involved in.

The farther we go, the more difficult it becomes to solve the problem of data exchange between these systems.

Additionally, if we take into consideration the cost of developing all interfaces (starting from laying out the lines, coordination of physical levels of signals etc. up to a unified language for communication with external information subsystems), then it becomes evident that only unique enterprises can solve information problems on their own.

It is more realistic to resolve these problems by a group of enterprises handling their common production problems within, for instance, the former branch ministries (nowadays, associations, as a rule) or zones of the country (districts, regions, republics). These regional information exchange networks of enterprises may be a basis for Integrated Systems of Regional Information Exchange support. These systems correspond to the concept of the "informatization of Russia" program which is being developed by the government of the country under difficult economy conditions.

The developers of the program have proposed not only a general concept but also "pilot" subsystems of various levels in the entire system. For example, the IPIAN (Institute for Information Transfer Problems, Russian Academy of Sciences) develops a version of local computational network means for the current conditions of Moscow region. The program incorporates the task of creating a "pilot" typical integrated system for a zone (a region, large cities).

While developing such system at TsAGI, we, due to financial limitations, plan carefully the stages of forming the organizational network within a regional one (developed jointly with VIMI, i.e., Russian Interbranch Information Interchange Institute) and gradually add new functions to these systems.

While the functional characteristics of the organizational system at all stages of its development are prescribed, we would minimize the expenses for the effort. Inversely, one may wish to maximize the functionality on the basis of the resources assigned.

Taking into consideration similar demands imposed on various organizational systems, we choose the means for implementation so as to unify them and minimize the expenses for the further development of an internet integrating the groups of organizational and regional networks.

As to a structure, an organizational system is a two-level system (Fig. 1) where the upper level comprises regional information centers linked between themselves and external information/communication systems by high-speed channels; and the lower, user's level comprises information centers of enterprises of the South-East of Moscow region and the subdivisions of the basic organization, i.e., TsAGI. Via specialized communication lines the organizational system is linked to an interregional information center of VIMI which exercises coordination and control in the integrated regional complex.

The regional information center software should be installed on a powerful high-speed computer with large volumes of an external memory, and extensive set of means for information exchange with the outer world (Fig. 2), and a complete set of means for interaction with organization information centers. Such an approach permits the user to significantly reduce the expenses for purchase of interfaces of various levels.

As a host computer for the regional centers, one can envisage ES EVM (Fig. 4) or western prototypes (such as IBM/370).

Advantages of this computer are as follows:

- availability of powerful multi-user operating systems (for example, VM/SP, MVS);
- availability of industrial data base management systems (SQL/DS, ADABAS, etc.),
- a perfect applications software (VTAM, TSAF, AVS, etc.),
- the architecture of this computer makes it possible to efficiently support and add diverse peripherals,
- the opportunity to incorporate a domestic inexpensive remote-access equipment,
- the external data storages can handle large amounts of information.

The most popular data-bases (domestic and foreign) are built, as a rule, on IBM/370 computers. Usually, the DBs are developed by professional programmers who know the network operation standards for DBs.

Systems based on IBM PC/386 or 486 have much poorer potentialities for handling the communication channels and peripheral devices. As well, the level of the DBs currently available is not adequate. However, such systems can essentially raise functionality of regional centers (Fig. 4).

Although the characteristics of an IBM PC do nc allow it to be a host computer, it can well be a workstation at an organization information center, see Fig. 3.

Software will include options for

- matching the interfaces of differing levels,
- accumulation and unified storage of information, and
- data retrieval.

With this, the end user of the center will be offered the possibility to utilize a simple set of software and hardware.

The regional and organizational centers will be a basis for information exchange subsystems:

1) major DBs for scientific, technical, and patent information, for international, State, and branch standards;

2) DBs receiving information from commercial networks (such as RELCOM, INFONET, "commercial channel", ITAR-TASS), radio and TV, and so on;

3) DBs of domestic and foreign organizations as to their products (to support the direct links between customers and manufacturers);

4) DBs with information from Exchanges; establishing an electronic Exchange, an multiregional dealer offices;

5) subsystems for surveying data from various DBs in order to facilitate decision-making at end users in commerce, management (for instance, the regional programmes for privatization);

6) e-mail subsystem connected with external (State-scale and foreign) data bases, maybe, through space communication networks (Fig. 4).

We are sure that TsAGI possesses all premises for implementation of the integrated information exchange system:

1) a reliable and powerful mainframe (the ES EVM) with considerable data storages;

2) a programmable communication controller based on microcomputers and PCs; it supports the communication function of the ES EVM of the regional center; the hardware is produced in small batches, complies with international and State standards, has been approved in a set of organizations (VIMI, MTsNTI, VNIIPAS, etc.), and enables the ES EVM to operate with international and State networks;

3) unique (for Russia) comprehensive DBs including scientific, technical and patent information, standards and practices, price estimation methods, etc.; the total amount is 10 gigabytes;

4) basic components of a regional network which include the above things, a specialized software and highly qualified developers who have a 10-year experience in the area, within a large research center;

5) both software and hardware are unified to a high extent, thus allowing us to assume these to be easily reproducible; a standardized basic configuration may be established and used in new regional centers.

The basic regional center can be formed in 6 to 8 months (duration depends on a configuration/type of links between regional and organizational centers). An estimated price of the project with

- the auxiliary equipment for a regional center and 8 organizational ones,
- a satellite communication station,
- the cost for incorporation in the networks, and
- the leasing of the communication channels

is some US $ 870,000.

Currently, TsAGI implements this project. The institute has bought (and plans to accomplish the assembly/installation work by the end of 1993) a satellite communication station ("Nauka") with a 7 m dia. antenna. The station offers reception/transmission of TV programs and up to 40 communication/radio channels through one of "Gorizont" communication satellites.

We have concluded the contracts for leasing the channels in the retransmitting satellite, acquired a license for commercial usage of communication channels, established a direct connection with international phone lines of the U.S.A. (via a teleport in New York).

If customers will so desire, we can sequentially implement regional systems, offer connection to computational and information-exchange networks, and improve/expand the configuration of the regional complex.

TsAGI would like to involve foreign specialists/companies experienced in this area and/or ready to make investment (equipment, finance, own communication options). The project is anticipated to help upgrade concepts in both the technology and management as to integration of regional, international and industry-owned networks. According to the "Russian Communication Regulations", networks and the relevant options may be owned by natural and artificial persons including the foreign ones. The amount of foreign non-governmental investment in Russian communication system obviously increases: US $ 120 million in 1992 and $ 350 million in 1993 (as expected). It would be clear that this sphere of activities is favourable.

Diagram 1

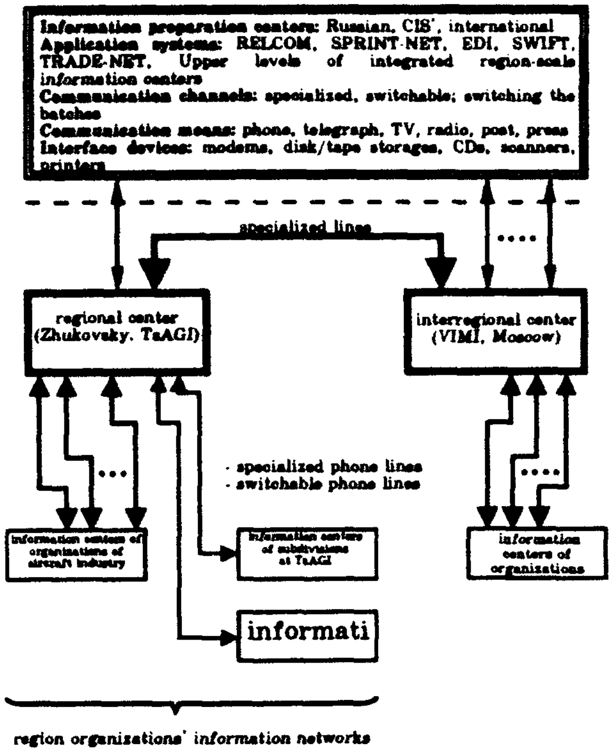


Diagram 2

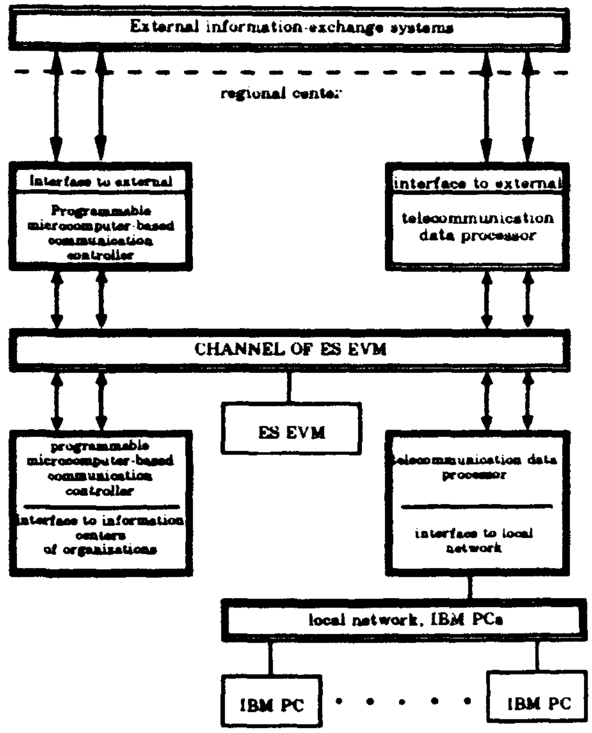


Diagram 3

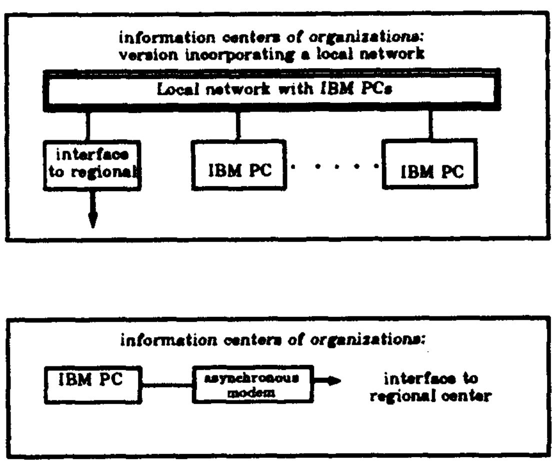
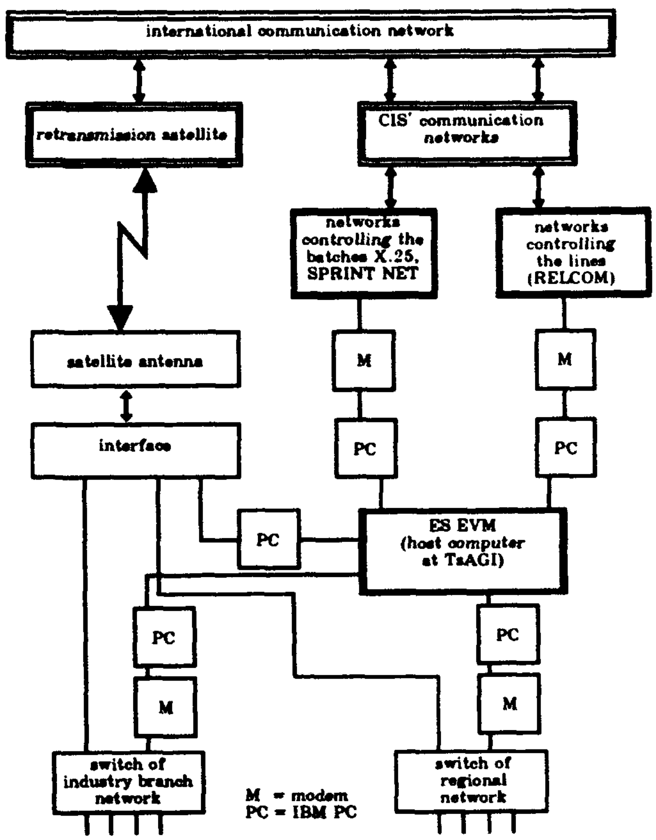


Diagram 4



ELECTRONIC DOCUMENT DELIVERY - TOWARDS THE VIRTUAL LIBRARY.

M. Brandreth and C. MacKeigan,
Canada Institute for Scientific and Technical Information (CISTI),
National Research Council of Canada,
Ottawa, Canada. K1A 0S2

ABSTRACT.

Reduced funding, global competition and technological change are forcing libraries and information centres to turn to new methods of providing services. Publishers of scientific and technical information expect printed journals to remain their primary product for several more years, despite the advent of electronic journals. Library collections will continue to be largely paper-based, but clients will demand much faster document delivery services from them. Many libraries must also maximise the investment in their collections by expanding their clientele. Electronic scanning of documents coupled with transmission over high speed, high capacity networks offers a potential solution to these problems. For more than a year, CISTI has been using proprietary imaging workstations to supply documents to one of its Branches. Much more flexibility is needed to reach the disparate receiving equipment used by a varied Canadian and international clientele. The paper describes experience with the Ariel Workstation, developed by the Research Libraries Group, and CISTI's own work towards a generic imaging workstation, able to transmit to a variety of receivers including identical scanning workstations, other workstations, facsimile machines and microcomputers with facsimile boards. Ability to rationalise library collections is seen as an important consequence.

1. INTRODUCTION.

This paper presents one aspect of the current uncertain world of libraries and information centres. It describes the effects of one new technology from the viewpoint of a high volume document supply operation faced with budgetary constraints and changing client demands. Although the experience reported has been gained mainly with equipment connected by dedicated lines or lower speed packet-switched networks, it is nevertheless relevant to the higher speeds of the Canadian network CANARIE and Internet, and leads us into some of the issues raised by networking. We will concentrate on current, practical aspects, leaving 'gee whiz' descriptions of what is certainly possible to later in the programme. For libraries that have to increase revenue in order to survive, the visions and achievements of research and development need to be put on a solid production basis.

CISTI, the Canada Institute for Scientific and Technical Information, possesses one of the largest collections of scientific and technical information (STI) in North America, if not the world. From our main library in Ottawa, we serve Canadians in industry, government, hospitals and universities. To serve the laboratories of our parent National Research Council we operate 15 Branches spread across some 5,000 kilometres from St. John's, Newfoundland to Victoria, British Colombia, the distance equivalent, for the benefit of our European participants, of London to Tashkent.

We will focus on the Document Delivery service, which on average supplies over 1000 photocopied documents a day, or nearly 2.5 million pages a year. About 80% of document orders are sent to us electronically (by ENVOY 100, the Canadian public messaging service, and CAN/DOC, CISTI's automated document ordering function), and we key in the rest. Order processing is quite well automated, including the identification and addition of library call-numbers to the order print-outs and the preparation of invoice data. On the other hand, the photocopies themselves are still produced by a manual photocopying process. Through work study, we have improved productivity by such means as flat bed machines, foot pedal operation, rapid scanning and workflow optimisation. Current photocopiers take about 2 seconds to scan a page and a further 2 seconds to return, during which time the next page is put in place. This labour intensive process is followed by slow delivery of the photocopy to the client via the Post Office or commercial courier services. At the client's request and for a premium charge, we send urgent deliveries by fax, but the machines generally use a single-page feed, hence need a prior photocopy step, and for some clients they give poor resolution and inferior quality.

For several years, we have, like many government organisations, had a static budget, which has been rapidly diminishing in real terms taking into account the escalation of journal subscription prices far beyond the inflation rate. In 1992, CISTI prepared a Strategic Plan, the development of which included a Delphi study. The Delphi experts diverged in opinion on the future of scientific and technical libraries, but the essential findings were clear and not surprising: networking, distributed processing, increasing application of the virtual library concept, and for smaller libraries 'just-in-time' services replacing 'just-in-case' collections. However, publishers particularly predicted that paper documents would probably be around until the turn of the century, as would the large depository libraries and document supply centres. Hence CISTI's large STI collection would remain the mainstay of our business for several more years, as the basis for new innovative services.

The Strategic Plan led to a business plan, in which we are challenged to reduce our dependence on government funding by dramatically increasing revenue (20% to 50% a year for the first 3 years).

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

Document Delivery being our prime source of revenue, we aim for a large increase in the already large number of small transactions, augmented by high volume business deals with brokers and research organisations, plus a move into the international market. This must be done by investing in new technology to increase productivity, reduce costs and provide better services to satisfy clients' increasing demands. The technology for document delivery becoming available during our planning suggested we should examine electronic scanning of documents and electronic transmission. Though existing telecommunications would suffice for some time, in the long run high speed networks would be commonly used to reach a far spread clientele, particularly CA*net and its successor CANARIE for Canadian clients and Internet for international clients.

2. ARIEL WORKSTATION.

In October 1992, a new CISTI Branch was set up to serve a new NRC biodiagnostics laboratory in Winnipeg. A Branch following the model of the existing ones would have had an on-site document collection and two or three staff; but our tight budget could no longer provide such resources. Instead, the new Branch had to rely on rapid on-demand service from Ottawa. Winnipeg being 2,000 kilometres distant, mail and courier services were deemed too slow, but the Ariel workstation had just become available and presented an opportunity for our first test of the virtual library concept.

The Ariel workstation was developed by the Research Libraries Group (RLG) in the US, and several university libraries were heavily involved in testing it. It comprises proprietary software running on specific off-the-shelf hardware. Each workstation consists of a scanner (Hewlett Packard Scanjet II or Panasonic), a PC-compatible microcomputer running the DOS operating system, and a Hewlett Packard II laser printer. In our case the telecommunications link was a 56 kilobit/second government-shared line using the TCP/IP protocol.

RLG's intention was to develop an *affordable* package for a *medium size* university interlibrary loans department using off-the-shelf equipment. Our experience has shown the concept to be sound. Ariel has proved very reliable. Clients find the output, printed on bond paper, to be of quite acceptable quality for both text and graphics at the high resolution setting of 400 dpi, which we almost always use. For half-tones the picture quality can be considerably improved by the electronic process of dithering, but at the expense of lower text quality and a slower scan. Ariel retains scanned images until a complete document has been compiled, and, to ease delivery, many documents can be assembled into a single transmission 'envelope' for each requester. Delivery, dependant on traffic on the line, may take as long as 24 hours, which the clients find adequate.

Nevertheless Ariel has proved to be unsuitable for our high volume operation, because of its slow scanning speed and limited user interface. These factors are manageable at the relatively low number of 30 items a day sent to Winnipeg, but would be intolerable for the several thousand items a day targeted for our total business. The HP ScanJet scanner supported by the software has a scanning time of 18 seconds per page (which dithering increases to 25 seconds). This equates to an average handling time per page of 40 seconds, compared to 4 seconds for photocopying. A document feeder has given only partial improvement. It needs single pages hence prior photocopying, and the software is not designed to support it. The loading limit set by the keyboard buffer is only 16 pages; the operator hits the Enter key 16 times and is then free for something else. The scanned documents cannot be viewed for editing purposes.

Being proprietary, the Ariel software is not supplied with source code or transmission protocol, hence it cannot be modified or emulated. This makes it a standalone system communicating only with another Ariel and not integrated with other library automation. This has not been a severe handicap when dealing with our Winnipeg Branch, but precludes application to thousands of users.

3. FAXON.

In 1991, CISTI signed an agreement with Faxon Research Services in Boston, MA to provide electronic copies of journal articles from our collection. The system was originally intended to be group IV fax based, but it was soon changed to imaging and transmission by file transfer. In Boston, Faxon converts the images and forwards them to its clients by fax. Electronic store and forward is another possibility.

The system, developed by Faxon during the course of the project, uses off-the-shelf hardware and custom-made software. The transmitting workstation consists of scanner and microcomputer. The main advantage lies in the scanner (Fujitsu 3096 with KOFAX interface compression/decompression board). This is comparable to a photocopier in having a scanning time of 3 seconds per page, with a 3 seconds return, conveniently the time the operator requires to position the next page. The scanner has a flat bed, eliminating the need for prior photocopying, and because we have control of the software, we have been able to rig a pedal to free the operator's hands. The resolution is usually set at 300 dpi, though 600 dpi is possible, and the quality of output is quite acceptable for text and graphics. Grey scales are used rather than dithering. After scanning, the document can be viewed, and during scanning single pages can be repeated to correct such things as poor positioning.

The software platform is the OS/2 operating system on an IBM token ring local area network with IBM LAN Server network manager in Ottawa and Banyon network operating system in Boston. These are different from CISTI standards, and so complicate the work of support and trouble shooting. The line to Boston is a dedicated 56 kilobit/second line and IBM's SDLC protocol is used to transfer files, which may consist of several documents. There are many links in the communications chain, decreasing the overall reliability significantly below that of individual components. For a high volume operation, the user interface is limited, and the software makes little use of library application protocols.

The past two years have proved to be a valuable learning experience, thoroughly confirming, as with Ariel, the concept of the virtual library. The principles can be applied widely, and are likely to set the pattern for other high volume agreements.

4. GENERIC IMAGING WORKSTATIONS.

These two experiences confirmed the immediate feasibility of electronic document delivery based on the scanning of printed documents from our collection. They also demonstrated the lack of commercial products meeting our requirement of a high volume/low cost installation serving a disparate group of clients. Suitably developed scanning technology instead of the labour intensive photocopying operation could help us meet our business plan target of improved productivity, particularly when scanned documents can be stored and used repeatedly. It is conceivable that eventually a document order could be received, processed and delivered to the client without human intervention.

Earlier this year we issued a Request for Proposals to develop and implement an electronic document delivery system (EDDS). The requirements document was very demanding but gave some latitude. For example, from our photocopying operation we set a target scanning time of 6 seconds per page consistent with our current workflow. However, we specified a longer total processing time instead of scanning time to permit novel approaches, such as taking a mobile scanner to the document on the shelf rather than going to the shelf and taking the document from it to a central scanner and back again. Our main requirements included a single workstation transmitting to a variety of receivers, ease of use, flexibility, and efficiency. We also looked for integration with our existing and future support systems, such as invoicing, tracking royalty payments, and client files, and eventually access to and retrieval of electronic full-text documents. In mid-September from the three companies responding, we selected an Ottawa-based company, Network Support Inc., which handles consulting, system delivery and support services for LANs, in association with PSS (Public Sector Systems), which has developed two relevant software packages for record and image processing.

5. ELECTRONIC DOCUMENT DELIVERY SYSTEM (EDDS) OVERVIEW.

The key features of the proposed EDDS are:

- it uses off-the-shelf hardware and software for imaging, document management and workflow;
- it will be built using client-server technology;
- it will help CISTI prepare for the future by offering the capability of storing documents as images;
- it is expandable, e.g., disks, memory, CPUs and server can be upgraded to support the optical disks needed for document storage.

Its principal software components are: RIMS (Recorded Information Management Software), DIMS (Document Image Manipulation Software) and FLOWMAN (a workflow application developed by Logical Software Solutions). All the proposed software is Windows based, giving a graphical user interface with point-and-click features.

The proposed hardware has several components:

- Servers - A Tricord Superserver configured with 2 CPUs (66MHz and 33MHz) for multi-processing and able to support 128MB of memory and over 50GB of disk drives. Two other servers will control the delivery of documents via both fax and other electronic means such as the Internet and X.400.

- Workstations - Scanning workstations consisting of Fujitsu 3096E scanners and 486DX 33MHz microcomputers with 8MB of memory and 130MB disk drive. Operators' monitors will be 17" for easy viewing of the images; supervisors' workstations will have the more common 14" monitors. An Intermec bar code reader will also be part of each scanning workstation.

- Printer - A single, fast printer, the ATI LC-6850, has been proposed to handle all printing requirements. It is a high performance laser printer rated at about 35 pages per minute for the EDDS application, which will be printing CCITT Group III or Group IV TIFF (Tagged Image File Format) images. It has three input trays with a 3,500 sheet capacity, can handle up to 11"x17" paper and can print on both sides of the page.

- LAN: A Novell LAN will link workstations distributed over several library floors.

- A software interface will handle data passing to and from our existing mainframe system.

Document orders received electronically will continue to receive some initial processing by the mainframe, e.g., automatic call numbering. To each order will be added appropriate client information taken from the client file, e.g., the client's address and preferred method for receiving documents - X.400 including text and images, TIFF formats, SGML (Standard Generalised Markup Language), fax, or manual mail delivery. Processed orders will be sent periodically from the mainframe to the EDDS, which will distribute them to particular sectors of particular stack floors, where "pick lists" will be printed out. For each order the bibliographic information needed to retrieve the item and a bar code corresponding to the order number will be printed. Staff will retrieve the requested items and take them to a workstation for scanning. The bar code will be wanded and the pages scanned. Staff will be able to view the document after scanning, verify the legibility, and, if necessary, delete or add pages.

The EDDS will change the status of the order as it proceeds through the workflow. When the status changes to "scanned" the delivery/printing operations will be initiated.

The system will automatically count the number of pages scanned for each order. This information, which currently is entered manually, will be sent with appropriate client information to the automated invoicing system. The data will also be collected for copyright reasons.

Other features include tracking the progress of the order and providing status information to clients on request, monitoring outstanding orders and if necessary redistributing them to ensure prompt processing, and reporting on process and queue times so that workflow may be adjusted for speed and efficiency.

6. NETWORKING CONSIDERATIONS.

As stated at the beginning, we have gained this experience with slow lines and relatively low volumes. Clearly with an expanded and international business, Internet will be used increasingly because clients will demand it. This raises several issues, some not nearly resolved. We will mention a few related to production operations.

By and large Internet links research and educational institutions and its funding takes that into account. Many of CISTI's clients are to be found in such institutions, but many are not. Canadian industry has a high proportion of small companies, many of which have no research departments and are not linked to CA*net. Hence, for each order we will have to use the transmission route selected by the particular client. In addition our charges include absorbed telecommunications costs from the commercial networks. As telecommunications for the Internet user appear virtually free, at least for the present, we may have to introduce differential pricing.

A tenet of document suppliers is reliability. Clients must be sure that their orders have reached us and that we will respond within a stated delay. We are proud of our record in this regard, and strive to help users with any problems in reaching us, whether the deficiencies lie in our systems or in the telecommunications. With Internet there is a major complication. Essentially nobody is responsible for managing Internet. Thus unlike the Canadian Datapac commercial network, which is clearly managed by Bell, there is no central point to call for help in diagnosing problems. We already have clients using custom-designed workstations and unsupported 'freeware', connecting to their institute networks and then to the nearest Internet node, and located on the other side of the continent. Thus there are many intermediate stages in the link to our computer, but no single organisation responsible for all of them. In such a situation it is virtually impossible to find out what went wrong if we fail to receive a client's order. Hence, we will not be able to guarantee any response to orders over Internet unless some sort of checking system is put in place for particular (high volume) users.

Standardisation of order formats is another problem. Currently clients place orders through a few defined systems, which, through prompts for original keying or transfer of records from our own databases, result in our receiving orders of a known structure. The EDIFACT (Electronic Data Interchange for Administration, Commerce and Transport) coding, part of the ILL protocol, delimits components of a document order. These codes are stripped out by our processing software (e.g. for automated call numbering) and we keep to the EDIFACT standard so only one stripping routine is needed. Any non-standard orders are keyed in by CISTI staff at a premium. In a sense, however, Internet may act against standardisation. Users will want to send us document orders comprising bibliographic references obtained in an unknown variety of ways - from online public access catalogues, from current awareness services, from online database searching, and by keying in data from printed bibliographies and other publications. Whether we should charge a premium for non-standard orders is an unresolved issue, likely to be influenced by market forecasts of volume and the perturbation to our workflow.

Another serious issue is copyright, which must be taken into account in any wide-scale use of the generic imaging workstation of the EDDS. Clients are demanding copyright-cleared documents, and CISTI is negotiating with copyright collectives and publishers; but much more clarification is needed in a very complex legal situation. As we understand the situation, transmission to store-and-forward devices is currently not allowed as it would permit multiple copying without control. Internet with WAIS, ARCHIE, gophers and as yet unthought-of means of accessing information sources could provide lawyers with perpetual business.

7. CONCLUSION.

CISTI's experience points the way to resource sharing among libraries over long distances. Large libraries and document suppliers will require something like CISTI's Electronic Document Delivery System for high volume operations where speed is essential together with the business features of high reliability and client information management. They are correspondingly expensive to develop and must be paid for out of the revenues they generate. Smaller operations and friendly exchanges between libraries will require fewer features and will be available at low cost, like Ariel. The technology, already far ahead of the legal and administrative questions, will force us to find answers quickly.

LA FOURNITURE DE DOCUMENTS ELECTRONIQUES EN EUROPE : L'EXPERIENCE DE L'INIST

Christian LUPOVICI
Directeur de la Production
INIST
2 allée du Parc de Brabois
54514 Vandoeuvre-lès-Nancy Cedex
FRANCE

INTRODUCTION

La création de l'Institut de l'Information Scientifique et Technique (INIST) en 1988, partait de la volonté du Ministère de la Recherche de développer, en France, un centre multidisciplinaire de fabrication de bases de données associé à un centre de fourniture de documents pour la diffusion de l'Information Scientifique et Technique, d'abord pour les chercheurs français, mais également au service des chercheurs européens et dans le monde.

Pour cela, il fut décidé de mettre en oeuvre les moyens technologiques les plus modernes.

A chaque stade de la production d'information, une technologie spécifique a été adaptée et a été intégrée dans le système général d'information, depuis l'acquisition des documents jusqu'à la fabrication des bases de données PASCAL pour la Science, Technologie et Médecine et FRANCIS pour les Sciences Humaines et Sociales, et à la fourniture de documents primaires.

La chaîne de production de l'INIST est divisée en 5 systèmes :
- le traitement des documents et leur gestion, effectué par un système GEAC 9 000 ;
- le catalogage analytique des articles de périodiques et la saisie des résumés effectués sur un réseau de stations de travail et sous-traité à une société privée ;
- la numérisation ;
- le stockage des documents sur support optique ;
- la gestion des commandes de la fourniture de documents.

1. LES COLLECTIONS NUMERISEES

1.1 L'historique

L'INIST a été membre de deux projets européens : TRANSDOC et ADONIS. De ces deux expériences, il en est résulté un projet opérationnel : le Système d'Archivage Numérique (SAN).

Du projet TRANSDOC (1984-86), l'INIST a retenu que le transfert par satellite des textes numérisés était difficile à mettre au point et coûteux, mais que les techniques de numérisation image et de stockage sur disque optique numérique étaient maîtrisables à un coût raisonnable.
Du projet ADONIS, l'INIST a retenu que la technique de numérisation a priori et intégralement des textes devait se gérer facilement en association avec la base de données.

1.2 Le projet SAN

L'INIST a choisi 1 800 périodiques (sur son fonds de 27 000 titres) pour être numérisés systématiquement. Ces titres ont été choisis parmi les titres les plus demandés en fourniture de documents et représentent donc les périodiques coeurs de la littérature internationale.

Si l'on considère que la demande de documents est, pour une moitié, extrêmement dispersée (sur les titres et les années) et pour l'autre moitié concentrée sur un peu plus de 5 000 titres, il était logique de prendre parmi ces derniers, les titres dont on sait d'avance que chaque fascicule est sollicité pour la fourniture de documents.

En dehors du défit technique que cela représentait en 1990, la numérisation permettait de résoudre simultanément plusieurs problèmes :
a) l'assurance d'une disponibilité permanente des documents, à la fois pour la fourniture de documents et pour l'analyse à la base de données ;
b) une garantie de qualité constante de la reproduction et une rapidité de transfert du document ;
c) un processus automatisé ;
d) l'espoir de réduire l'espace de stockage à terme.

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

1.3 La technique employée

La numérisation est effectuée en mode image selon les normes recommandées par le groupe international sur l'échange de documents électroniques (GEDI) : une résolution de 300 bpi, l'organisation des fichiers en format TIFF et la compression des images en groupe IV CCITT.

La société JOUVE qui effectue la numérisation des 90 fascicules par jour, trie les documents en fonction des problèmes de qualité :

- 30 % des documents qui contiennent beaucoup de photographies sont traités systématiquement en "mode mixte" (technique proche de la norme JPEG) permettant de simuler une trame et convertir en groupe IV CCITT ;
- 70 % des documents peuvent être numérisés de façon classique pour économiser de la place. Seules les pages d'images posant problème sont reprises en "mode mixte".

Les images sont copiées sur des disques optiques numériques ATG double face qui contiennent 70 000 pages. Les disques sont chargés sur un Juke-box (CYGNET) qui contient 131 disques, muni de 2 lecteurs de disques (ATG 60001). Le système est géré par un ordinateur SUN 4/370 (sous ORACLE/UNIX) qui communique avec le système de fourniture de documents sur un IBM 9121 par l'intermédiaire d'un réseau local Ethernet.

Trois micro-ordinateurs permettent la transmission des documents sur les réseaux de télécommunication en fax groupe IV ou groupe III.

De la réception de la demande à l'INIST à la réception du document chez l'utilisateur, quelques minutes suffisent, et sans intervention humaine.

2. LA GESTION DES COMMANDES ET LA FOURNITURE DE DOCUMENTS

L'INIST traite environ 600 000 demandes de fourniture de documents par an, activité en progression d'environ 8 % par an.

Pour assurer ce service avec des délais courts (2 heures, 24 heures ou 48 heures selon le niveau de service demandé), l'INIST a informatisé la gestion des commandes.

2.1 Les différents moyens de commande

2.1.1 Les commandes sous forme électronique

Les clients de l'INIST peuvent utiliser différents moyens de commande :

- la messagerie des serveurs (QUESTEL, ESA et bientôt DIALOG) ;
- les systèmes de prêt entre bibliothèques tels que le PEB en France ou OCLC ;
- la saisie directe sur le système INIST par terminal vidéotex (Minitel en France) ou VT 100. Dans ce cas, l'utilisateur a accès en ligne au catalogue des documents (titres fascicules et articles de périodiques) de l'INIST et le système remplit automatiquement l'écran de commande lorsque le document a été identifié.

L'INIST a également développé des procédures de transfert de fichiers pour les clients les plus importants.

L'accès via INTERNET est possible techniquement mais le formatage des données doit faire l'objet d'un accord avec l'INIST pour que les demandes soient exploitées automatiquement.

2.1.2 Les commandes sous forme traditionnelle

L'INIST reçoit encore beaucoup de commandes sous forme non électronique :

- des formulaires papier ;
- des fax ;
- des coups de téléphone.

Toutes les informations sont, dans ces cas, saisies dans le système pour que toutes les commandes soient gérées électroniquement.

2.2 La fourniture de documents

Le système trie les commandes selon la cote en magasin des documents et adresse les commandes aux bons étages de magasin où elles sont imprimées pour traitement.

Le Système d'Archivage Numérique (SAN) est le magasin prioritaire. Toutes les demandes portant sur des documents numérisés lui sont adressées. Ce n'est qu'en cas de panne qu'un reroutage dans les magasins est prévu.

Pour les documents numérisés, le SAN traite les commandes par lot, toutes les deux heures. C'est l'équilibre actuel entre la nécessité de grouper les demandes par adresse physique sur les disques et la nécessité d'une réponse rapide au client.

Pour tout problème de référence incomplète ou ambiguë, le système ou les magasiniers adressent l'enregistrement au fichier d'un service spécialisé dans l'identification et la localisation des documents.

L'INIST travaille en coopération avec 100 bibliothèques de recours pour obtenir les documents qu'il ne possède pas. La localisation de certains de ces documents est faite dans le fichier INIST pour permettre une orientation externe automatique.

La saisie des éléments de fabrication sert d'indication de fin de traitement.

CONCLUSIONS

La création d'un système de gestion électronique des commandes offre énormément d'avantages :
- la gestion des commandes est précise et la régulation des flux plus facile ;
- la fabrication de statistiques est extrêmement précieuse pour notre politique de développement des collections et pour le choix des titres à numériser.
L'INIST fournit, au Centre Français d'Exploitation du Droit de Copie (CFC), deux fois par an les statistiques précises des documents fournis, classés par éditeur.
Le système de commande associé au système de stockage électronique doit permettre de fournir plus rapidement, plus de documents et sans intervention humaine de la commande à la réception (par fax).
Cependant, le système montre plusieurs inconvénients :
- la place prise par la forme image est relativement importante, surtout pour les documents comportant des photographies ;
- l'accès aux disques optiques est plus lent que l'accès aux disques magnétiques, et pour des volumes importants, cette différence est sensible puisqu'il faut pratiquement une journée de 10 heures pour traiter le volume théorique maximum actuel.
- le coût de la numérisation à partir de la forme papier est important, surtout si l'on considère les problèmes de qualité à résoudre quotidiennement.
Il faut ajouter que la statistique nous montre qu'un article est rarement demandé plusieurs fois et que beaucoup d'articles ne sont jamais demandés, même dans le cas de titres de périodiques très demandés en fourniture de documents.
L'intérêt de cette expérience est de nous permettre de nous familiariser avec ces technologies.
Cependant l'INIST étudie déjà la possibilité de gérer non plus des images de pages numérisées, mais le texte intégral fourni par les éditeurs eux-mêmes, dans sa forme électronique d'origine en mode caractère selon la norme SGML (Standard Generalized Mark-up Language). Nous savons déjà gérer les enregistrements bibliographiques avec le résumé en SGML.

Le problème est ici plus lié à la capacité des éditeurs à moderniser leur chaîne de fabrication et à leur volonté de trouver un accord de copyright (globalement ou séparément) dans le cadre de la diffusion des documents électroniques à la demande.

Electronic Journals

by

Cliff McKnight

HUSAT Research Institute
Loughborough University of Technology
Elms Grove, Loughborough LE11 1RG
UK

Introduction

In 1981 a Royal Society study of the scientific information system in the UK described the electronic journal as "perhaps the most radical innovation in prospect for the primary literature". Twelve years and as many projects later, has the electronic journal gone any way towards realising its potential? In order to answer this question, it is necessary to consider what the electronic journal has to offer. Or, to put it another way, what is wrong with the paper journals?

The common criticisms of the paper journals are that they are often not on the shelf (mis-shelved, gone for binding, being used by someone else); they are hard to search through; they take up too much space; and there are too many of them to keep up with.

In contrast, electronic journals are always there (assuming the network hasn't crashed), easy to search (but can flood you with hits unless you are careful in the way you conduct your search) and occupy much less space. In addition, various electronic 'filters' can be used to ensure that users receive notification of relevant journals — the 'selective dissemination of information' or SDI.

These are the commonly attributed general advantages of electronic journals. However, they do not seem yet to have been sufficient to establish the electronic journal firmly in the scholarly communication chain. In order to see why this is the case, some of the electronic journal experiments will now be considered. Finally, suggestions will be made regarding problems remaining to be tackled.

BLEND

The BLEND project (Shackel, 1982; 1991) aimed to investigate the feasibility of supporting the entire scholarly communication process — from authoring and submitting, through refereeing and editing, to publishing — via computer. A central mainframe was used and the various participants in the process communicated through this machine, with the resulting four issues of the journal *Computer Human Factors* being stored on it.

Although it is often assumed that the electronic medium would speed up the process of publication, this is not necessarily the case. Indeed, Shackel (1991) reported a median publication time of 'just over 32 weeks' (i.e. about 8 months) for articles in *Computer Human Factors* and it is possible that this could be attributed to a 'novelty effect'. What seems more likely is that the electronic medium more easily supports comment and dialogue than the paper medium, as will be seen in some of the more recent experiments described below.

In at least one respect, *Computer Human Factors* proved superior to a paper journal. Although each article was 'read-only' once archived, there was space allocated for comments to be entered on each article and these comments could then be seen by subsequent readers. Since the articles' authors were also part of the 'electronic community' this meant that they too could read and respond to the comments. The resulting dialogue created much more of a feeling of 'live' research than is possible in the paper medium where the time between submission and publication is often over a year.

While successful in this respect, *Computer Human Factors* was not without its problems, mostly due to the technology of the day. Users accessed the system via a remote terminal either over the Joint Academic Network (JANet) or the Public Switched Telephone Network (PSTN) using fairly low transmission rates. Typically, terminals would be located in the university's computer centre (which could well be in a different building to the academic's office) and would frequently be in use by other users. This is in contrast to the present position where most academics have some sort of computer on their desk (Shackel, 1990) and in many cases it is connected into the campus Local Area Network.

Movement through the articles was slow and screens were poor. Not surprisingly, therefore, users preferred to print copies of articles that they wanted to read in their entirety. The articles themselves were restricted to plain ASCII text and 'typewriter graphics'.

QUARTET

Project QUARTET aimed to investigate the implications of information technology for the scholarly communication process. It was therefore somewhat wider than BLEND, being concerned with a broad spectrum of communication activities including electronic mail, computer based

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

conferencing, electronic document delivery, desktop publishing and electronic publishing (Tuck *et al.*, 1990). As part of Project QUARTET, colleagues and I designed and built the world's first hypertext electronic journal, HyperBIT[1]. This was intended not as a replacement for the library archive copy of the journal but rather as an electronic version of the personal subscription. Our design was based on the results of various earlier studies by us of journal usage (e.g., Dillon, Richardson and McKnight, 1988) and as such specifically addressed the issue of user requirements. Hence, browsing through author/title lists at either the issue or volume level was supported, as was searching the entire contents of the journal. Each article was structured using the Guide™ hypertext system and cross-references in articles were made into active hypertext links, allowing the reader to move quickly and easily between articles. (A more complete description of the design is given in McKnight, Dillon and Richardson, 1991.)

HyperBIT offered the user several advantages over the paper version. For example, it was always available on the desktop (whereas the paper version might well have been borrowed by a colleague). The entire contents of the journal could be searched in order to locate, say, all articles which mentioned 'screen' or referred to work by 'Eason'. The ability to move between related articles using the hypertext links was also advantageous, as was a pop-up window facility which provided instant access to the bibliographic details of references without leaving the text. This facility was provided on the basis of observations of many users who would keep a finger permanently in the References section of the article when using the paper version, turning to the section when they encountered a reference in the text and then returning to the text. In this sense it provided an 'electronic finger'.

The chief disadvantage of HyperBIT compared to the paper version concerned graphics. Although the Macintosh system on which it was implemented allowed the display of quite sophisticated graphics, animation and sound, the screen resolution was far lower than the resolution available to produce the average paper journal. Typical screen resolution is 72 dots per inch (dpi) whereas a typical typesetting machine has a resolution of 1200 dpi — roughly 16 times better. In many cases this is not a real problem — the line art typical of diagrams and graphs presents no difficulty. However, the display of half-tone photographs, for example, presents a little more difficulty.

LISTSERV

The models of the electronic journal employed by BLEND and QUARTET could be described as 'centralised' and 'distributed' respectively. However, in recent years a 'mixed' model has arisen based on the LISTSERV software. This name is an abbreviation of 'list server' which gives some insight into how the system works. In a typical system, a central computer holds a list of subscribers; when a new issue is available, the system sends subscribers a 'contents page' and abstracts via email. Subscribers can then request articles by sending an email message to the server, with the articles being automatically delivered as email by the software.

Although the concept of 'issue' is still used, the issue itself is effectively unbundled since subscribers can request single articles. However, the contents pages and abstracts can be stored for future reference and searching, and articles can be retrieved at any time on demand. Such a system makes effective use of the network 'bandwidth' since only requested articles are transmitted. How many academics could honestly say they are interested in every article of every issue of every journal they receive? Even when the journals are on their shelves, they don't remember what is in them.[2]

Unfortunately, in order to reach a wide audience, the LISTSERV journals are transmitted in a form which makes few assumptions about the type of computer which will be used by the recipient. In practice, this means that they are usually limited to plain ASCII text with fixed line lengths. Hence, the appeal to the lowest common denominator is still as real today as it was for the BLEND project.

Also like BLEND, though, the listserv journals have found that they can support discussion of articles. For example, subscribers to the peer refereed electronic journal *PostModern Culture* (*PMC*) can also choose to subscribe to a discussion list called *PMC-Talk*. Indeed the discussion list provides a forum not only for comment on and debate about the articles in the parent list, it also provides a forum for discussion of the broader issues of postmodernism in general. A further facility offered by PMC is PMC-MOO, a "real-time, text-based virtual reality environment in which you can interact with other subscribers of the journal and participate in live conferences" (PMC-list, 1993).

CORE

The CORE (Chemistry Online Retrieval Experiment) project's aim is to deliver a large majority of the journal literature needed by one academic area in electronic form to workstations in

[1] The journal used was Behaviour and Information Technology (BIT), published by Taylor and Francis whom we gratefully acknowledge for allowing its use.

[2] Shackel (1985), for example, reports on an electronic search of a references and abstracts database: "...33 [references] were subsequently used in the preparation of the written chapter. Of these 33, 16 were already known to me, but 17 were new, highly relevant references...In almost all cases the relevant journals were on my bookshelves."

a library and terminals on the desks of academics. Articles are held in both text and bit-map forms and a variety of interface options are being investigated. For example, Landauer *et al.* (1993) report on experiments comparing performance on five different tasks using a SuperBook interface (which allows browsing and provides many hypertext-like features), a PixLook interface (which combines a sophisticated document retrieval engine with bitmap page images) and paper. Not surprisingly, the results suggest that there is no one 'best' interface. Rather, particular tasks are supported to a greater or lesser extent by each interface; as the task changes, so the optimum form of interface will change.

The CORE project represents a collaboration among five institutions: the Cornell University Albert Mann Library houses and administers the experiment; the American Chemical Society is providing ASCII and microfilm versions of the last 10 years of 20 journals; the Chemical Abstracts Service provides electronic versions of their hierarchical indexing scheme tagged to all of these articles; the Online Computer Library Center (OCLC) is contributing expertise in large database storage, access and search techniques; and BellCoRe (where Landauer and colleagues work) is contributing expertise on text and graphic conversion and transmission as well as developing prototype user interfaces. At the time of writing, the project is still running and we await its results with interest.

TULIP

The TULIP (The University Licensing Program) project, a three-year project scheduled to run to the end of 1995, aims to test the feasibility of networked delivery and use of journals. Elsevier Science Publishers are making electronic versions of 42 of its materials science journals available to the 15 colleges and universities (including MIT, Harvard, Carnegie Mellon, Cornell and Princeton) that are participating in the experiment. Each university is providing its own hardware and access and retrieval software. Hence, Elsevier are simply providing a database which Engineering Information, acting as Internet host, archive and customise for each university. The aim is to provide 'as much local autonomy as possible' (Elsevier, 1992) so that a variety of options can be explored and evaluated. The project will examine the economic, organisational and technical issues involved in the electronic transmission of journals as well as considering user issues.

The journals are currently stored as bit-maps and are distributed with index files and a 'dirty ASCII' file. This latter can be searched but not displayed. It is produced by scanning the journals and using OCR software to recover the text. Many people have expressed surprise that it is more cost effective for Elsevier to re-scan and OCR pages which have already existed in electronic form. However, it is likely that this is a temporary expedient in order to get the project off the ground — it was already delayed for some months beyond its scheduled start date. As typesetters move towards a standard format, so it will be easier to produce a 'clean' ASCII file directly. As with CORE, results from TULIP are awaited with interest.

OJCCT

The American Association for the Advancement of Science (AAAS) and the Online Computer Library Center (OCLC) have also launched an electronic journal, *The Online Journal of Current Clinical Trials*. This was due to be launched in April 1992 but was beset by technical problems and eventually started some months later. In addition to the technical problems facing the project, Wilson (1992) reports that "the AAAS must persuade authors to submit high-quality papers in a new medium that may prove to be largely ethereal." Indeed, the number of papers on the system does not seem to be reaching the expected level — at the end of February 1993 there were only 31 articles on the system, most of which had formed the launch batch. Only one new article was added during December 1992. Although this problem will obviously decrease as the number of quality electronic journals increases and authors become more confident of placing their work in the electronic domain, it seems that the situation is still not much different from that experienced by the early EIES project (Sheridan *et al*, 1981).

The project has tried to steer clear of dependence on a particular system or interface and to this end three routes are available to the articles: via OCLC, via CompuServe and via the Internet. Paradoxically, although the system was designed to be accessed directly by readers, it seems to have received a more enthusiastic reception from the information profession than from end users. Although it was not originally intended to publish papers in printed form, it was announced shortly after the launch date that the *Lancet* would publish abridged versions of the electronic texts. However, even this does not seem to have the desired encouraging effect on authors.

MSMSE

Lest it seems that currently only the USA is interested in the electronic journal, I would like finally to mention two British projects. The first involves the Institute of Physics Publishing and Loughborough University with support from SCONUL. This project, funded by the British Library Research and Development Department (as were BLEND and QUARTET), is looking at a variety of economic, technical and user issues involved in the distribution of an electronic version of a paper journal, *Modelling and Simulation in Materials Science and Engineering*, to participating libraries (Pullinger and Meadows, 1993).

This project has many similarities to the TULIP project: it seeks to involve both publishers and libraries; it is allowing participating libraries to specify the format in which they receive the electronic version of the journal; and it is distributing electronic versions of existing paper journals. This is in contrast to, say, the list-server journals which effectively bypass the publisher and library (or at least the formal representatives of these bodies) and impose a format on the distributed journal.

The Institute of Physics Publishing also led a consortium of nine publishers in a small project testing journal distribution and usage over SuperJANET, the new British high speed academic network currently undergoing pilot testing. This project aimed to produce a demonstration system and was scheduled to run from January to April 1993 (IOPP, 1993) so results are awaited with interest.

CAJUN

The second British project to start recently is the CD-ROM Acrobat Journals Using Networks (CAJUN) project. This project involves two journals already in existence, Wiley's *Electronic Publishing: origination, dissemination and design* (*EP-odd*) and Chapman and Hall's *Optical and Quantum Electronics* (*OQE*). Dissemination will be both on CD-ROM and over the network.

The 'Acrobat' referred to in the project title is Adobe Acrobat™ which is "a family of products that work together to enable document communications" (Adobe, 1992). The basis of Acrobat is the Portable Document Format (PDF) which is PostScript based but also allows additional document features such as annotations, hypertext links and miniature thumbnail views of each page.

Like PostScript, PDF is device and resolution independent and Acrobat viewers (the applications necessary to read, navigate and print PDF documents) are available for Macintosh, Windows, DOS and UNIX platforms. Adobe intends to publish PDF as an open standard.

In the same way that the TULIP and MSMSE projects are important because they explore the role of the publisher and library in the electronic journal, so the CAJUN project is noteworthy since it involves a major commercial software house — indeed, one which has already been responsible for producing the *de facto* standard page description language, PostScript.

Problems remaining

It is clear that part of the continuing need for electronic journal experiments is based on the fact that the underlying technology keeps changing. Faced with a different set of possibilities, it is important to investigate the advantages and disadvantages of each. If the type of target machine is known, quite sophisticated electronic journals are now possible, although the 'lowest common denominator' still seems to be ASCII text.

If we accept that the technology already exists to support electronic journals and that they could alleviate some of the problems associated with paper journals, it remains to consider the shortcomings of electronic journals. We began by asking 'What's wrong with paper?' but now we should consider its advantages since it may be technology's inability to mirror such advantages that is constraining the development of electronic journals.

One of the most commonly cited advantages of the paper journal is its portability. A survey by Simpson (1988) suggested that many academics like to read while on trains or at home rather than in their office or the library. A journal on CD-ROM could be carried easily between office and home but would require equivalent equipment at both places. However, there are certainly portable electronic books being developed and it may well be that the portable electronic journal will follow behind. The current growth in laptop computers and the recent arrival of 'notepad' computers, combined with a storage medium such as the smart-card, could well support portable electronic journals.

Scholars also point out that the paper journals are easy to browse and they often mention the serendipitous discovery of other articles in the paper journals. Such browsing and serendipity can be supported in the electronic domain, as demonstrated by HyperBIT but even in that case they were not as easy to perform as with a paper journal. We pick up an issue of a journal off a desk, flick through it and then drop it back on the desk all in a matter of seconds.

Although systems like HyperBIT allowed a reasonable level of graphics, this was achieved by making the system machine specific. Displaying the same graphics on different systems is still difficult — file formats do exist which allow this, but few software packages other than dedicated viewing programs can deal with such formats and movement between different systems usually requires modifying the file header information. Furthermore, the resolution of typical screens is not yet sufficient to allow the use of such graphic items as photographs — again, such systems do exist but they are far outside the purchasing power of most users.

Those electronic journals which are distributed over the network must also recognise the fact that they are most readily accessible to the academic market. There are many researchers located in industrial research laboratories who have no access to this network. In America at least, an increasing number of companies are connecting into the network, but in Britain the take-up rate is very slow in the industrial and commercial sectors. Furthermore, it must be recognised that any electronic journal excludes a large number of users of the paper journal system — those users in countries that do not yet have a stable network or even an established computer base. Clearly, the hope is that such countries will develop a computing infrastructure eventually. I would not argue that we should not develop electronic journals because they presently exclude such countries, but we must recognise that access for a large number of potential users is currently impossible.

Although it is not something which technology can provide, an important element missing from the electronic journal is its acceptance in the academic hierarchy. One of the fears expressed regarding electronic journals was that the ease of distribution over networks would lead to a lowering of quality. It is recognised that electronic journals must be refereed in the same way as the paper journals in order to maintain quality, yet even when this is done there is still a reluctance to publish in them, as OJCCT has discovered. The reason for this is that publishing serves more purposes than simple information dissemination in academia — careers are built on publications, tenure can be won and lost in the journals. It is likely that electronic journals will be acceptable eventually, but in the meantime it is still a gamble publishing in them.

The question of copyright control is of particular concern for publishers. Although the paper medium is relatively easy to copy using a photocopier, the resulting copy is of inferior quality to the original. In the electronic domain, copying is not only easy and fast but also the resulting copy *is identical to* the original. If I receive an article over the network, it takes me no more than a few key-presses to forward a copy of the entire article to someone else. This means that either methods of electronic copy protection must be developed or the concept of copyright must be reconsidered.

In the paper domain, if I order a journal I either have to pay for it myself (a personal subscription) or I must have agreement that the department or university library will pay. If I receive a listserv journal, however, it is not clear who pays. Certainly there are costs involved — there is no such thing as a 'free lunch'! — but they are costs which are largely transparent to the user. The storage and staff costs are typically met by the host institution, usually a university and in this respect we may be witnessing a return to the situation in which universities were also publishing houses. My access to the network is paid for as part of the general funding for computing within the university and I don't receive a bill. Hence, it may well prove necessary to develop new costing models for the production and distribution of journals in the electronic domain and several of the current projects are actively considering such issues.

Conclusion

While the paper journal will be with us for many years yet, it is clear that the electronic journal is here to stay. However, many of the problems remaining to be solved are of a psychological, social and political rather than technological nature. What is also clear is that it is not sufficient for a journal simply to be electronic in order to guarantee its success. We need to be able to perform *at least* the range of tasks which we can perform with the paper journal and preferably *more*. We need *usable* electronic journals and this means designing them in a user-centred way with the clear emphasis on user needs rather than technological possibilities.

References

Adobe (1992) Adobe™ Acrobat™ Products and Technology: an overview. Adobe Systems Incorporated, November.

Dillon, A., Richardson, J. and McKnight, C. (1988) Towards the development of a full text, searchable database: implications from a study of journal usage. *British Journal of Academic Librarianship*, 3(1), 37–48.

Elsevier Science Publishers (1992) TULIP: an irregular update. N° 1, November.

Institute of Physics Publishing (1993) Electronic journals to be pioneered on SuperJANET. Press release, 21 January.

Landauer, T., Egan, D., Remde, J., Lesk, M., Lochbaum, C. and Ketchum, D. (1993) *Enhancing the usability of text through* computer delivery and formative evaluation: the SuperBook project. In C. McKnight, A. Dillon and J. Richardson (eds.) *Hypertext: A Psychological Perspective*. Chichester: Ellis Horwood. 71–136.

McKnight, C., Dillon, A. and Richardson, J. (1991) *Hypertext in Context*.. Cambridge: Cambridge University Press.

PMC-list (1993) Postmodern Culture announces PMC-MOO. Message posted to subscribers to pmc-list@listserv.ncsu.edu, 21 May.

Pullinger, D.J. and Meadows, A.J. (1993) Distribution of an electronic version of a scholarly journal from publisher to libraries. Paper distributed to exchange of experience meeting, Royal Society, 26 February.

Royal Society (1981) *A Study of the Scientific Information System in the United Kingdom*. BL R&D Report 5626. London: Royal Society.

Shackel, B. (1982) Plans and initial progress with BLEND — an electronic network communication experiment. *International Journal of Man-Machine Studies*, 17, 225–233.

Shackel, B. (1985) Using RAAJ for real. *Journal of Librarianship*, 17(3), 200–204.

Shackel, B. (1990) Information exchange within the research community. In M. Feeney and K. Merry (eds.) *Information Technology and the Research Process*. Sevenoaks: Bowker-Saur. 147–162.

Shackel, B. (1991) BLEND-9: Overview and Appraisal. British Library Research Paper 82. London: The British Library.

Sheridan, T., Senders, J., Moray, N., Stoklosa, J., Guillaume, J. and Makepeace, D. (1981) Experimentation with a Multi-disciplinary Teleconference and Electronic Journal on Mental Workload. Report to National Science Foundation, Division of Science Information Access Improvement, June.

Simpson, A. (1988) Academic journal usage. *British Journal of Academic Librarianship*, 3(1), 25–36.

Tuck, B., McKnight, C., Hayet, M. and Archer, D. (1990) *Project Quartet*. LIR Report N° 76. London: The British Library.

Wilson, D.L. (1992) Major scholarly publisher to test electronic transmission of journals. *The Chronicle of Higher Education*, June 3, A17 and A20.

BULLETIN BOARDS, ELECTRONIC MAIL, CONFERENCING. CURRENT USE BY SCIENTISTS AND ENGINEERS; EFFECTS ON LIBRARIES AND INFORMATION CENTRES - DO THEY HAVE A ROLE ?

Víctor Castelo
Communications Manager

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC)
Centro Técnico de Informática
Pinar 19
28006 Madrid
Spain

E-mail: castelo@cti.csic.es

1. ABSTRACT

This paper introduces the present scenario of networking, the different systems of exchange information person to person, the way to access the information and to know where the information is.

A summary of the more utilized applications is made, some of them have appeared in the last years as the result of a normal evolution: Bulletin Boards, Gopher, WAIS, World-Wide Web, E-mail, Conferencing, Directory service, etc. . All these applications represent an added value to networking allowing the user (scientist, engineer, librarian, etc.) to communicate with others on their same area of interest. Additionally the user can obtain information about how to connect to servers and get items, programs, data, etc., and to follow conferences or make questions in a forum.

As an example of an important application of distributed directory the Paradise project is discused, it yields some tools in order to look for information on institutions, people and other objects.

Taking all these considerations into account, it is intended to make an analysis about the current use of these tools by engineers and scientists. The role that should play the Libraries and Information Centres in this context is also discussed.

2. KEYWORDS

Electronic networks; Engineers; Scientists; Information Systems; Libraries; Information Centres.

3. PRESENT SITUATION

In the last few years the development of Networking and Computing has offered the possibility for accessing to a great part of the huge quantity of information available on electronic media to a high number of people.

New communication networks are appearing (e.g. EuropaNET [1]) and the already existent ones (e.g. JANET [2]) are growing up, also the bandwidths constantly increase on the links. The logical networks over these infrastructures experience an enormous growth, particularly the Internet. This network is growing up in such a rhythm that a change of the address system for the computers is being considered. Otherwise the domain of numbers will be exhausted in few years [3].

The new communications technology is being widespread over many countries and the use of ISDN (Integrated Service Digital Network) is a reality. In some instances Broadband-ISDN is being applied as the new products need more and more memory, CPU power and communications (bandwidths).

ATM (Asynchronous Transfer Mode) and Frame Relay are the new technologies more employed today for fast packet switching with the bandwiths actually used.

Also appearing are applications of audio and video conference taking advantage of the new data transmision systems that could be interesting in some cases, e.g., teleconferencing and teaching.

The information is there, on the machines, on the networks, but one must search to it. A series of applications have interaction with the user in order to allow him access the information, as for making a connection to a database, interchange E-mail with a colleague or capture a file containing programs from a remote machine.

But when the possible sources of information multiply and a great dispersion exists, the classical applications

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

have revealed insufficient. We need new applications capable of guiding the user, locate the information and also make easier the task of connecting to a remote database, with a user-friendly interface and client-server environment.

The more used applications today are products that allow to locate where the information is, driving the searching process. They are like bulletin boards. The use of information systems with local data and pointers to access to other distant sources is constantly increasing. Examples are Gopher, WAIS, World-Wide Web (WWW), etc.

In the majority of cases the software is public domain, easily obtained from the network file servers.

Let us assume that one has an Internet with only a TELNET (remote login in Internet) or a FTP (File Transfer Protocol). In this case we might have much information but the "navigation" to find it would be very difficult.

With new programs like Gopher we could jump from system to system by selecting within a menu or clicking the mouse to search and retrieve the information.

4. THE APPLICATIONS

We briefly overview the new and classical applications more used at present on the networks.

4.1 Gopher [4][5]

Gopher is a system of menus in a tree-structure that yields direct information (stored on its own files), and allows the access to other information systems like other Gophers, WAIS, WWW, etc. . Also transfer files by means of FTP or provide interactive connections to databases in other machines. The original design was done at the University of Minnesota.

Normaly it is used in a client-server structure. It is available in a multitude of platforms and operating systems, as other products of this kind There are versions for Macintosh, MS/DOS, OS/2, X-Windows, VAX/VMS, VM/CMS, etc. .

By means of Gopher it is posible to reach objects from different nodes. Always in a very intuitive way, following a menu without worryng neither about lots of addresses nor technical commands. For example, in order to acces a database we only select a point of the menu and the system executes the connection commands for us.

When we have located an information, it is possible either to store it in a file of our own client or (if we use a remote client) request sending it by E-mail to our address.

The capability of access to multimedia objects like sounds or images are only available with some clients.

The majority of the Gophers allow to generate marks (bookmarks) in order to speed up the access to any point of a menu even in an low level without going diving into the tree.

4.2 WAIS (Wide Area Information Server) [4][5]

WAIS is a very used system designed by Thinking Machines Corp. to search and retrieve distributed information.

From the technical point of view it employs the standard Z39.50 (ANSI) [6], which together with SR (Search and Retrieve, ISO 10162 and ISO 10163) are the protocols more used to enable the request of services between information systems.

In order to do a search we have many databases available. We should first select some of them and later choose the keywords to execute the search. For example in XWAIS (WAIS client for X-Windows) we must only fill the "In sources" and the "Tell me about" fields. After the search we obtain various documents with a score showing how appropiate each one is. The best has a score of 1000 and the others proportionally less.

The final retrieve of the primary document can be carried out by request to our local client or by means of an indication to the remote system to send us a copy by E-mail.

4.3 WWW (World-Wide Web) [4][5]

World-Wide Web is a information system based on hypertext, the first design was done at CERN (Switzerland). The information in our screen is related to other documents with links and the reader can jump (if he wants) to the explanation of one particular word in other document, which can have other possible links to new documents and so on.

The links are marked for reference with numbers within square brackets, italics, highlighting or with coloured words.

At any moment a list of the "visited" documents can be obtained. It is possible to return quickly to any of them.

The information is distributed and can be found in various systems. With appropriate tools it is possible to store our own information in order to participate in the searchs.

4.4 E-mail

The electronic mail is a very useful application, perhaps the most employed. The normal use is to exchange correspondence between persons. There are also many systems with automatic reply to an E-mail.

E-mail is a powerful aplication over many networks using different protocols (SMTP, X.400, EARN-BITNET, etc.). There are systems called gateways to link the different environments, the user does not need to know the way followed by his message to the recipient.

The distribution lists by electronic mail (LISTSERV in EARN-BITNET [5]) are based on an automatic system with an address associated to a file containing the E-mail addresses of a group of persons. When the list receives a message it automaticaly sends a copy to all the subscribers.

To be subscribed and "signoff" on a list it is necessary to send a message to LISTSERV or a similar application indicating the action and the name of the list.

The distribution lists are also employed to distribute electronic magazines.

Other interesting applications of the E-mail are the file servers or databases where it is possible to obtain a file (e.g. FILESERVER in EARN-BITNET) or search and retrieve an information (e.g. ASTRA in EARN-BITNET and ARCHIE in Internet). The request is made by means of a message and you receive the answer by E-mail.

4.5 News

NEWS [4] is a service available in various networks and very similar to LISTSERV. Many lists of LISTSERV are included in the NEWS, but in this case the users do not receive copies of the messages; instead the messages (about the same discussion) are stored in a computer and distributed to other interested computers using a special protocol called NNTP (Network News Transport Protocol).

The person who requests to read a conference establishes a connection to the nearest computer with NEWS service or uses a client on his own PC linked to a News server.

The user could follow (browsing) an enormous number of conferences on the most varied topics, from bonsais to Nuclear Physics.

The reader can participate in a discussion including his own comments or sending a particular answer by E-mail to the person who had sent a previous message with a question.

4.6 Directory

Sometimes it is necessary to locate a person or an institution and get basic data like their telephone or postal address, etc. During the last years different initiatives have existed, but the project Paradise [7] is leading the course in a worldwide development, being quite promising.

It follows the standard X.500 (OSI) of distributed directory. The records stored in the DSAs (Directory System Agent) are rapidly increasing over the world.

The information has a hierarchical structure , similar to a tree in a computer file directory (UNIX or MS-DOS). The top level is the world and in descending order we can find the organizations, the departments and finally the persons.

The information is multimedia and images and sound could be contained, like people photos or maps showing where to locate one organization within a city.

The data related to persons can include a great number of fields like activity, postal address, fax, telephone, E-mail and some curious items as favourite drink, prefered method for contact, photo, etc.

Several user agents exist (DUA: Directory User Agent) for multiple platforms with different degrees of friendship and capacity to perform a search.

5. THE APPLICATIONS USAGE

Taking into acount the state of the art on network technology along with the number of applications now available, a user who has access to some of the networks spread out worldwide could get the available information in a relatively comfortable way and on his own.

One could read his E-mail, reply some letters and look for the address of a person in other country with X.500. Afterwards, one could make a connection to a database, search a book and obtain a file with a conference abstract via FTP.

The workstation or PC are the neccesary tools opening a window to a world of networks full of information. There, the user satisfies his needs using the friendly applications, described above.

Sometimes the value of the information [6] obtained in a search carried out personally (whenever satisfactory), without a third party, is very high since it is obtained in a record time.

The engineer and the scientist need a type of information essentially similar, but the final products are different [8].

The Scientist has plenty of contacts with his colleagues, he uses very much E-mail [9] and constitutes part of "invisible colleges". The engineers who work in an enterprise have a relationship very close within their own enterprise for obvious reasons.

Nowadays many groups of persons are working in different countries using services like E-mail or distribution lists avoiding the high cost of meetings.

In some cases the scientist has a "call back" service to use his Computer Centre, workstation and the wide-area communications free of charge from home.

Can the user himself, at this moment, navigate by the networks searching the information?. The answer is YES, but then, what should the Information Centres and the Libraries do? and what role should they play at present?

6. ROLE OF THE LIBRARIES AND INFORMATION

Obviously the Information Centres and the Libraries have a lot of things to do with these new technologies. First of all they should familiarize with them and break certain barriers that exist for different reasons.

The person who uses these technologies is not an expert, and although the majority of the current applications are easy to use, it would be interesting to have a short training organized by the Information Centre [10], perhaps in collaboration with the Computer Centre in some cases.

Our experience with scientists is that they attend in great number a seminar about some of these applications that they think are of great utility and frecuently use, if the time required for the seminar is about a few hours.

The Information Centre must be involved to a great extent in the design of the new systems, their maintenance and their management.

The existence of a good collaboration between the Information Centre and the Library with the Computer Centre is always very interesting [11].

The update of the information is very important. This requires the set up of an information system from the informatic point of view. However the daily work maintaining the information is a task for Libraries and Information Centres. They must also perform the editorial control [12] in the case of electronic magazines distributed by networks.

There are numerous bulletin boards, including some specialized with information about Libraries and useful for librarians, where this type of institutions have been collaborating in a very active way, like BUBL [13] in JANET (UK), which is accessible from NISS (National Information on Software and Services) [14]. This opens the door to a great number of very interesting applications.

In the information systems you can obtain the addresses of OPACs (Open Public Access Catalog) [12] and, using systems like Gopher, it is possible to get directly an online connection.

The Libraries have many things to do in the processes of standardization, compilation and testing of availability, with all the network products. They should make a great effort so that the search systems will be easy to use [10].

Although the user can "navigate" by himself, in some cases the Information Centre staff could make a search with better success because a better search strategy could be setup due to the specialist's professionality.

The sources are well-known for an expert and the access to particular systems with specialized information could go beyond the knowledge of the user. However, the Information Centre or the Library, by means of their contacts with other Information Centres are the way to know some special sources.

In any case, fee-paying databases, where it is necessary to make a subscription and a password is given for accessing, must be used, obviously, under the supervision of the Information Centre, even though they were reachable through free networks.

Also the Information Centre could make personal services, and supply, by means of specialized documentalists, a data analysis that introduces a great added value to the information [8].

7. CONCLUSION

The increase of the information sources and the accessibility by networks is allowing the appearance of new enhanced applications in order to navigate easily in the search of information.

The scientists and engineers make an increasing use of these techniques in a personal way, employing their own workstation or PC to obtain almost every information needed.

Nevertheless, the Information Centres and the Libraries have an active role. They should participate in the design, development and the maintenance of the applications and the information on the networks. They should be very related to their Computer Centers and should finally integrate into the new documental chain, using the present technologies, from selection and acquisitions up to the retrieval of the primary document.

8. ACKNOWLEGMENTS

My acknowlegement for their ideas and for the material supplied to the Centro Técnico de Informática of CSIC, specialy to its Information Centre, the Unidad de Coordinación de Bibliotecas (CBIC) of CSIC and the Centro de Información y Documentación Científica (CINDOC) of CSIC.

9. REFERENCES

[1] Dai, D., "EuropaNET- The implications of a multi-protocol backbone", in 4th Joint European Networking Conference, May 10-13, 1993 (RARE 1993, Trondheim).

[2] Cooper, R., "SuperJANET. The Gestation of a High Performance National Research Network", in 4th Joint European Networking Conference, May 10-13, 1993 (RARE 1993, Trondheim).

[3] IETF (Internet Engineering Task Force) 26th meeting, 1993.

[4] Krol, E., "The Whole Internet", O'Reilly & Associates, 1993, Inc.,ISBN 1-56592-025-2.

[5] EARN, "Guide to Networks Resource Tools", 1993

[6] Husby, O., "SR A protocol for open library and information systems" in Information Networking 93, London, May 1993, ISBN 0.88736-936-7.

[7] Goodman, D., "El servicio de Directorio de COSINE (PARADISE)", Boletin de RedIRIS, 13, September 1991.

[8] Altuna, B., "El valor de la informacion desde la perspectiva del usuario y el papel de las bibliotecas y unidades de informacion en la empresa", Documentación Científica, Jul-Sep 1992, ISSN 0210-0614.

[9] Bishop, A., "Electronic networking for engineers: research from a user perspective", Computers Networks and ISDN Systems, Vol 25, November 1992 (RARE 1992, Innsbruck).

[10] Ríos, Y., "Catálogos en linea de acceso público: selección bibliográfica", Documentación Científica, Apr-Jun 1991, ISSN0210-0614.

[11] Huizer, E.; Abbema, J.; Heijne, M.; Verschuren, T.; "Reaching out to the end user", in 4th Joint European Networking Conference, May 10-13, 1993 (RARE 1993, Trondheim).

[12] Kelly, P., "Academic Libraries: "Towards Commonwealth coalitions" ", in Networking and the future of Libraries, April 2-5, 1992, Paper 8.

[13] Meadows, J., "Education and training" ,in Networking and the future of the Libraries, April 2-5, 1992. Paper17.

[14] Żedlewski, E., "National Information on Software and Services (NISS)" in Information Networking 93, London, May 1993, ISBN 0-88736-936-7.

TRAINING & OPERATIONS SUPPORT SYSTEM (TOPS)

by

Guy Langlois and Ian Gordon
CAE Electronics Limited
PO Box 1800
Saint Laurent
Quebec H4L 4X4
Canada

Abstract

This paper describes a potential training and operations support system and discusses the enabling technologies that make it possible – high speed global networks, powerful computers, and an object oriented data base management system (OODBMS). It illustrates a real world example of the system using a scenario involving an aircraft maintenance technician.

Introduction

Time and technology march inexorably onward. For those involved in the maintenance of complex systems, keeping up with technology is essential. We rely on systematic approaches to information storage, retrieval and training methods to manage and maintain the technologies we devise. Yet the sheer volume of information creates delays and bottlenecks, making timely and cost-effective access difficult.

However, the problems created by technology are often best resolved by technology. It has become feasible to discuss a future where information is as available a commodity as the air we breath. We can easily imagine databases linked over global high-speed computer networks; and sophisticated software to package, store, index and disseminate huge and varied chunks of information. It is possible to foresee a time when a person engaged in an important and complex task might, via a simple device, gain access to up-to-the-moment information allowing him or her to complete the task in minutes instead of hours or days.

This paper describes such a potential system. TOPS (Training & Operations Support System) provides real-time, on demand access to distributed training and support information. This information may include on-line electronic support documentation, schematics, illustrations, video, audio, animation sequences, three-D, layered models, simulation based training systems, reference materials, maintenance procedures, vendor and public libraries, historical performance data bases, test environments or even live guidance from other remotely situated individuals.

TOPS also provides a basic infrastructure of system services that support the timely delivery of media rich content to the information users. These services include electronic mail, video-conferencing, object tagging, automated configuration control, dependency tracking, data traceability, as well as services for system administration and management.

TOPS reduces the restrictions placed upon training by cost and geography. It improves the quality, consistency and availability of training and operations support information, thereby reducing costs and providing a vehicle for coping with future change.

Sample Application

Let's look at an example scenario where an aircraft maintenance technician named "John" is performing a preventive maintenance procedure on an aircraft hydraulic system. He logs onto to a TOPS work station (or a hand held device) and sees that the maintenance task involves a procedure he had not performed recently (the aircraft he is working on has a new instrumentation configuration). Before starting the procedure, John decides to refresh his memory and directs TOPS to display a list of relative training materials and support information. John selects a multi-media tutorial overview from the list and in an instant he is flipping through an electronic book, reviewing animations and exploring layered 3D models. John now wants to review the interaction of this particular hydraulic system with other aircraft systems. He sees from his search hit list that simulation based training material is available for the aircraft type and selects it from the list. In a few moments John is reviewing system schematics, invoking malfunctions and analyzing the corresponding effects on other systems.

John feels comfortable about what has to be done and proceeds to complete the remaining steps in the procedure. During system tests John notices a discrepancy in actual system readings with those published in the test procedures. John investigates the discrepancy but the situation remains unchanged. Through TOPS, John requests a real-time video conference with the on-line maintenance support group. Based on keywords supplied by John, TOPS searches for and initiates contact with a qualified expert. Within minutes a live picture of Sam appears. Sam is an expert in the hydraulic systems John is working with, who happens to be physically located in the aircraft manufacturer's engineering facility on another continent. John and Sam review the situation, procedures performed, system status, and published test result tables. Together they decide to compare aircraft systems status with that of a simulated test environment of the specific aircraft configuration. John connects the TOPS workstation to the aircraft maintenance port. The aircraft feeds system status parameters into the simulated test environment. Status readings in the test environment are identical to those of the actual aircraft but Sam realizes through the simulation models that a valve pressure

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

reading is inconsistent and suspects the valve to be defective. Sam directs John to replace the valve and then retest the system. As it turns out, all tests are successful and system readings are normal. John invokes another video conference session with Sam to tell him the good news and a thought he had about improving the procedure.

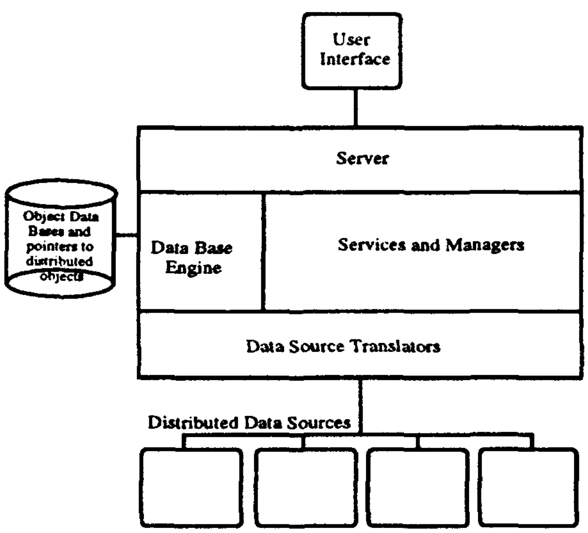
In a few moments Sam reappears and John reviews the test results with him. John suggests that the procedure be expanded to include verification of related valve pressure readings. Sam agrees, writes up a short note, records a brief audio message and sends it off through TOPS to the originator of the maintenance procedure.

Sixty minutes have passed, an enormous amount of information has been exchanged, an important task was performed efficiently and effectively, and the entire process was improved along the way. How long would it take to replay this scenario in today's world, with traditional training and support methods?

System Overview

TOPS consists of three major components as shown in the following figure and described below:

- front end or client workstations with an intelligent user interface that allow users to easily access information and seamlessly collaborate with each other.
- data sources including training media libraries, technical libraries, historical performance libraries, simulation data bases, test environments, etc.
- servers with powerful data base management systems responsible for storing information and pointers to distributed information as well as translating user requests into queries that data sources can understand (in order to accommodate "legacy" information systems). Services such as video conferencing, search and query tools, and real-time delivery of time sensitive information are all also provided.



The user interface is similar to current on-line services in that information categories are displayed to a user, e-mail and information searching facilities are available, and file transfers are possible. It is different from these systems in that the information categories represent integrated training and operations support data located in many geographically separated locations, the quantity of information exchange is higher, data types may include sound, audio, video, animation, 3D models, and high fidelity simulations – all delivered to users in real-time. The user interface allows for seamless access to training media data bases, procedure libraries, on-line documentation, historical maintenance data bases, document libraries, image libraries and simulation model libraries.

Between the interface and the data sources lie distributed servers. It is here where data base objects or pointers to distributed data base objects reside along with interpretation and translation software, dependency information and services that ensure the timely delivery of time-sensitive data to end users.

Several enabling technologies make TOPS possible, including high speed networks to deliver the required volume of information in a timely manner, object oriented data base management systems (OODBMS) that can manage complex information and its dependencies across distributed systems, high performance hardware (becoming a commodity) and powerful (object oriented) software. The underlying OODBMS technology required by TOPS is discussed below.

Data Base Engine

Our implementation of TOPS would use an object oriented technology for managing information across distributed systems. The OODBMS engine is based on a client server architecture, can be fully distributed, supports up to four giga-objects within the same database and four giga-databases on the same system. Classes are objects themselves and the database supports *containers* which can hold up to four billion pointers to other database objects.

Pre-defined Objects, Classes and Services

In addition to common OODBMS services for doing SQL-like transactions, for manipulating objects and classes, for updating schemata and for integrating new tools to the database, the OODBMS includes a set of pre-defined objects, classes and services for grouping objects and classes into re-usable software components, compressing object data, viewing, listening and interacting with objects and for performing configuration control on objects. The OODBMS maintains knowledge of complex dependencies between each object whether it be video, audio, or data.

System Management and Configuration Control

Objects can be grouped into configurations, and configurations into baselines. Each configuration and baseline can have a "role based" password protection scheme and separate historical files. Baselines may reflect the contents of the database at a time of proven performance or specific time (for example: approved training program or maintenance procedures). Configurations may be listed, compared, copied, saved, restored, renamed or deleted. Object data can be shared by multiple configurations. A new history record is registered whenever additions, changes or deletions are made to each configuration. It is also possible to get insight into the impact on other database objects for a given transaction via pre-defined classes and services for manipulating complex objects. For example: when entering a complex object under configuration control (i.e. one that may involve several tightly coupled objects or files), the OODBMS services automatically knows what type of object it is and what related objects to look for.

Browsing and Queries

A Browser is required for viewing, listening and interacting with objects within the context of a given Configuration or Baseline. By invoking OODBMS services, the Browser interfaces to the computer vendor multi-media tools to drive audio or video output as well as most popular documentation tools in a manner which is transparent to the user. The browser offers complete traceability from an abstract object name to one or many data files which constitute any version of the complex object. For example, the user can use a configuration management tool to group several documents, digitized video clips and simulation models under one topic, lets say: "B747-400 Hydraulics Overview". Another user can use the Browser to *PLAY* the "B747-400 Hydraulics Overview" without having to know that the course is built from several Objects. The user would simply select the 747-400 Training Courses Configuration , *LIST* the available courses and *PLAY* the "B747-400 Hydraulics Overview". The browser can also be utilized to "BROWSE" through the database searching for particular instances of objects or object data.

Tailoring - Configuration Control

A Schema Editor is needed to add, delete and modify schemata in the database and for developing new services for tools and objects. The Schema Editor should allow system administrators to define new classes, change the inheritance of existing classes as well as rename, delete or modify classes and their attributes. New schemata and associated database services can be imported to an existing database while other users are accessing it. The database engine will automatically verify that the newly imported schemata created by the Schema Editor will not cause the existing tools to potentially corrupt the database.

Summary

Training is today's technological world is on-going. Technology is constantly changing and product life cycles keep getting shorter. Maintaining adequate knowledge and skill levels among personnel is becoming more difficult and expensive. Nonetheless, for many companies it is often imperative that operations and maintenance personnel thoroughly understand the technologies they work with on a daily basis.

TOPS provides real-time, on-demand information to users at their place of work. Information may include simulation based training materials, simulation based job aids, interactive multi-media presentations, instructions, tutorials, books (e.g. maintenance procedures, video conferencing, lectures, seminars), live on-line access to diagnostic and systems experts and high fidelity simulation aids.

The system relies on three maturing technologies to be effective: high speed networks, object oriented data base management systems, and powerful hardware and software. These technologies are close enough to reality that planning for this type of system is becoming appropriate.

WIDE AREA INFORMATION SERVERS: AN EXECUTIVE INFORMATION SYSTEM FOR UNSTRUCTURED FILES

by

Brewster Kahle and Harry Morris
WAIS Inc.
1040 Noel Drive
Menlo Park
CA 94025
United States
brewster@wais.com
morris@wais.com

Franklin Davis[1], Thomas Erickson[2], Clare Hart[3], Robin Palmer[4]

Abstract

In this paper we present a corporate information system for untrained users to search gigabytes of unformatted data using quasi-natural language and relevance feedback queries. The data can reside on distributed servers anywhere on a wide area network giving the users access to personal, corporate, and published information from a single interface. Effective queries can be turned into profiles, allowing the system to automatically alert the user when new data is available.

The system was tested by twenty executive users located in 6 cities. Our primary goal in building the system was to determine if the technology and infrastructure existed to make end-user searching of unstructured information profitable. We found that effective search and user interface technologies for end-users are available, but network technologies are still a limiting cost factor.

As a result of the experiment we are continuing the development of the system. This paper describes the overall system architecture, the implemented subset, and the lessons learned.

1. Introduction

Systems that allow corporate executives to access personal, corporate, and published information such as memos, reports, manuals, and news are new in the field of information management. The first integrated systems are just now coming on the market. They exploit networking, online mass storage, and end-user search systems; each of these has existed for some time, but their combination and integration has not been available for the corporate environment.

Commercial systems exist in each of the personal, corporate, and published data areas with different levels of user friendliness. ON Location™, for instance, allows easy content based retrieval of personal files on a Macintosh, while Lotus Megellean™ performs a similar function on a PC. Verity's Topic™ system allows for searching of LAN-based (usually corporate) archives but primarily for a trained user community. Dialog, Dow Jones, and Mead Data are major online providers of published information, but again the majority of their users are professionals in the field of information retrieval (such as corporate librarians).

Academic systems have also been developed for some of these applications. The Information Lens project (Malone, 1986) revolves around structured electronic mail to help in automatic organization and retrieval of business information. Project Mercury (Ginther-Webster, 1990) is a remote library searching system that uses a client-server model. The Smart system (Salton, 1971) is an information retrieval system that embodies many different searching strategies. The SuperBook project (Egan, 1989) is working on user interfaces for information systems concentrating on the scientific user. Each of these systems is breaking new ground, but there is still no complete solution for the business executive wishing to search diverse information sources.

The Wide Area Information Servers (WAIS, pronounced "ways") system was constructed to test the acceptability of an integrated search system directly targeted at executives (Kahle, 1989). The companies participating in the project offered expertise in different parts of the problem: Dow Jones, with its business information sources; Thinking Machines, with its high-end information retrieval engines; Apple, with its user interface background; and KPMG Peat Marwick, with its information-hungry user base. Through this project, we wanted to find out if the wide area information retrieval market could incorporate users other than those trained searchers who are familiar with a variety of query languages and databases.

In the WAIS project we used a general architecture and built a small implementation to test the feasibility of an integrated information retrieval system for corporate end users. This paper is a report on the overall architecture, the various implementations, and the lessons learned from this work.

The WAIS Architecture

The WAIS system took advantage of available technology to make a system which could then be tested on corporate executives to determine user acceptability. The system was

1. Thinking Machines Corporation (fad@think.com)
2. Apple Computer (thomas@apple.com)
3. Dow Jones & Co.
4. KPMG Peat Marwick (palmer.r@applelink.apple.com)

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

composed of: clients, servers, and the protocol which connects them. The information servers were Connection Machine systems, running a parallel signature-based search algorithm (Stanfill, 1986). The cross-country network connected several LANs with leased lines running AppleTalk and TCP, and carrying a variation on the Z39.50 application protocol. The clients ran on Macintoshes. This section will describe the overall architecture, and the next section will describe what exactly was implemented and used during the experiment.

The WAIS architecture had the following goals:

- accessible to novice users - little or no training should be required in order to perform effective searches.
- remotely accessible - the servers must be accessible over a variety of networks.
- uniform interface - a variety of databases, whether personal, corporate or published, must be accessible from the same user interface.
- automatic alerting - it must be easy to create profiles for background searching
- scalable - the system must scale in number of servers, size of servers, and intelligence of servers.
- security - individuals and groups should be able to maintain control of who accesses their data.
- pricing model - a variety of information pricing structures, from per-minute charges, to subscriptions, must be supported.
- multimedia - the system must support the retrieval of any file format.

Many of these goals were achieved, while others, such as pricing model experimentation, were left unresolved.

In a client-server system, the client program is the user interface, the server does the searching and retrieval of documents based on indices, and the protocol is used to transmit the queries and responses. The client and server are isolated from each other through the protocol so that they can be physically distant and interchangeable. Any client which is capable of translating a user's request into the standard protocol can be used in the system. Likewise, any server capable of answering a request encoded in the protocol can be used. In order to promote the development of both clients and servers, the protocol specification is in the public domain, as is its initial implementation.

On the client side, searches are formulated as quasi-natural language questions. The client application then formats the query for the WAIS protocol, and transmits it over a network to a server. The server receives the transmission, translates the received packet into its own query language, and searches for documents satisfying the query. The ranked list of relevant documents are then encoded in the protocol, and transmitted back to the client. At this point, the servers do not "understand" the quasi-natural language question posed by the user in any sense that a human would, but they use the words and phrases in the question to find documents that use those terms. The client decodes the response, and displays the results. Documents of interest to the user can then be retrieved from the server.

2.1 *Searching*

We modeled the searching strategy on the interactive process people use when talking with a reference librarian. The library scenario is one where the patron approaches a librarian or researcher with a description of needed information. The librarian might ask a few background questions, and then draw from appropriate sources to provide an initial selection of articles, reports, and references. The patron sorts through this selection to find the most pertinent documents. With feedback from these trials, the researcher can refine the materials and even continue to supply the patron with a flow of information as it becomes available. Monitoring which articles were useful can help the researcher provide appropriate information in the future.

The WAIS system uses a similar means of interaction: the user states a question in unrestricted natural language to a set of sources, and a set of document descriptions is retrieved (see figure 1). The server assigns each document a score, based on how closely the words in the document matched the question (see figure 2). The user can examine any of the documents, print them, or save them for future use (see figure 3). If the initial response is incomplete or somehow insufficient, the user can refine the question by stating it differently.

Once a good document is found, the user may say "I want more like this one" by marking the retrieved documents as being "relevant" to the question at hand, and then re-running the search (see figure 4). This method of query refinement is called "relevance feedback" (Salton, 1983). The server uses the marked documents to attempt to find others which are similar to them. In the present WAIS server, "similar" documents are ones which share a large number of statistically significant words and phrases. This brute force method works surprisingly well with large collections of documents (Stanfill 1986; 1991).

2.2 *A Common Protocol for Information Retrieval*

One of the most far-reaching aspects of this project was the development of an open protocol. The four companies involved jointly specified a standard protocol for information retrieval by extending an existing public standard, Z39.50-1988 (NISO, 1988). We choose this public standard rather than inventing one ourselves since it was close to what we needed and it could help us keep the protocol from being regarded as proprietary.

The use of an open and versatile protocol can foster hardware independence and competition. This not only provides for a much wider base of users, it allows the system to evolve over time as hardware technology progresses. For example, the protocol provides for the transmission of audio and video as well as text, even though at present most personal computers are unable to handle them. However, they are free to ignore pictures and sound returned in response to questions, and to display and retrieve only text, if that is all they are capable of processing. Higher-end platforms are free to exploit their greater processing power and network bandwidth.

Z39.50 is a general attribute-based Boolean search protocol intended to run over the OSI stack. It was designed for search and retrieval of bibliographic (MARC) records in libraries. As such, its structure allows easy access to traditional Boolean search systems such as STAIRS (Salton, 1983).

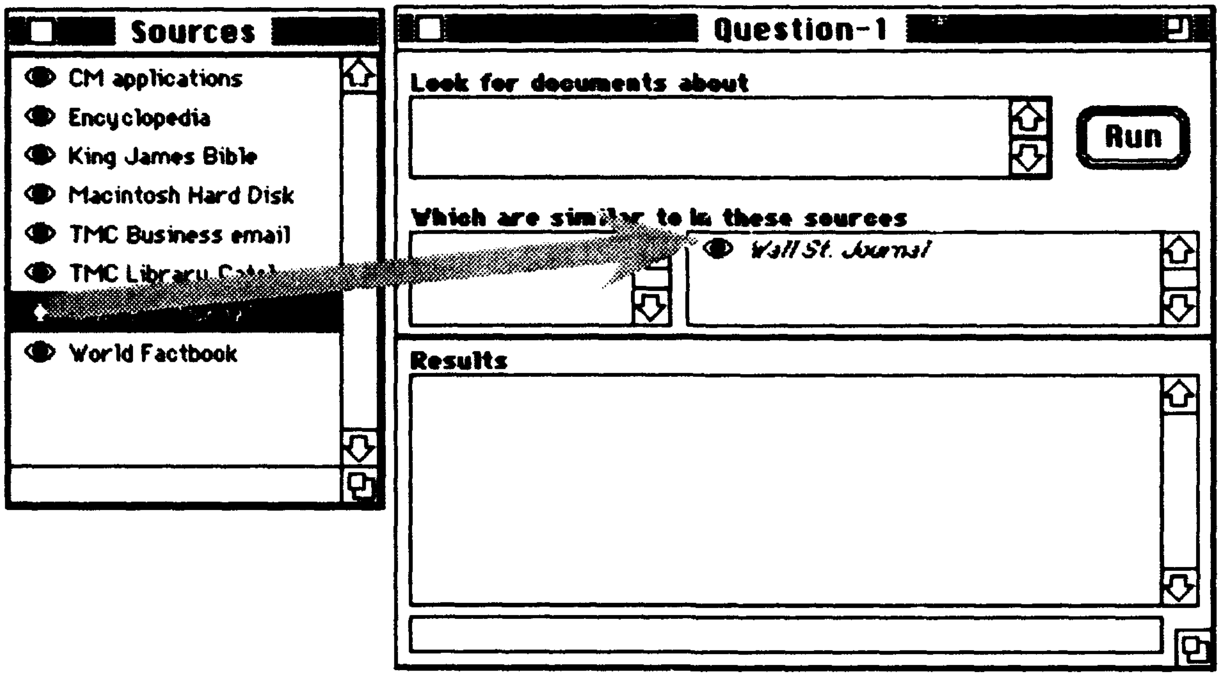


Figure 1: Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

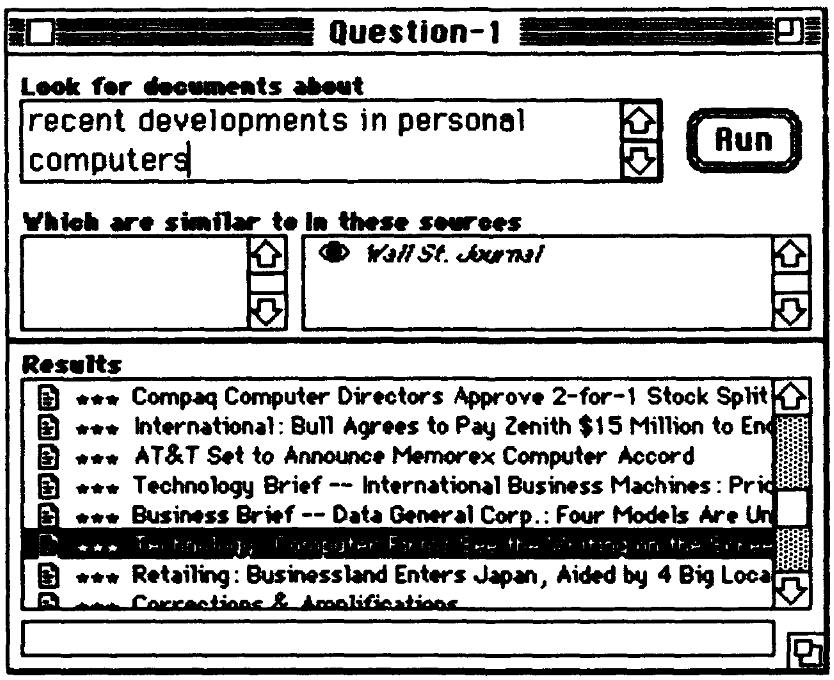


Figure 2: When a query is run, headlines of documents matching the query are displayed.

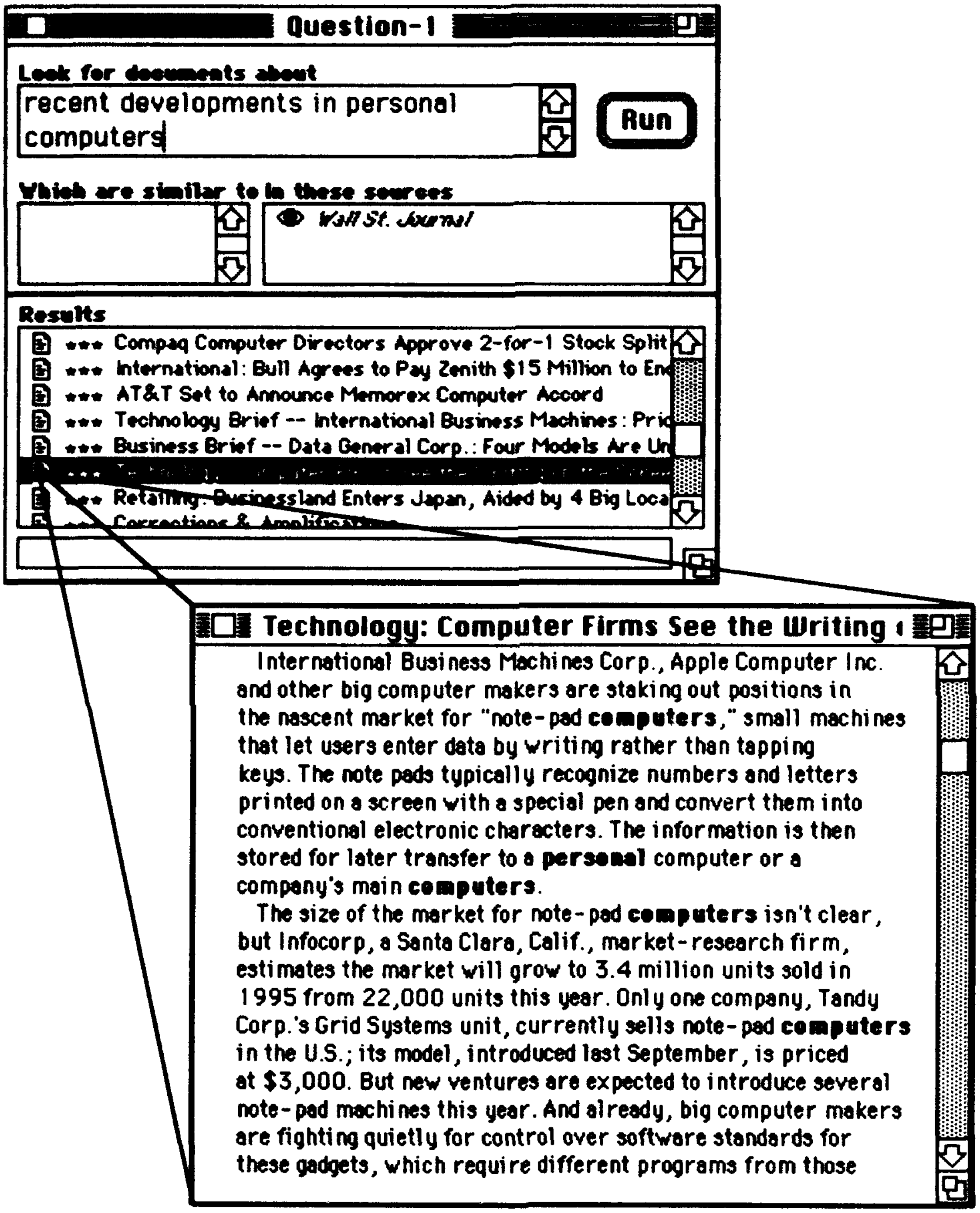


Figure 3: With the mouse, the user double clicks on any resulting document to retrieve it. The document can contain graphics.

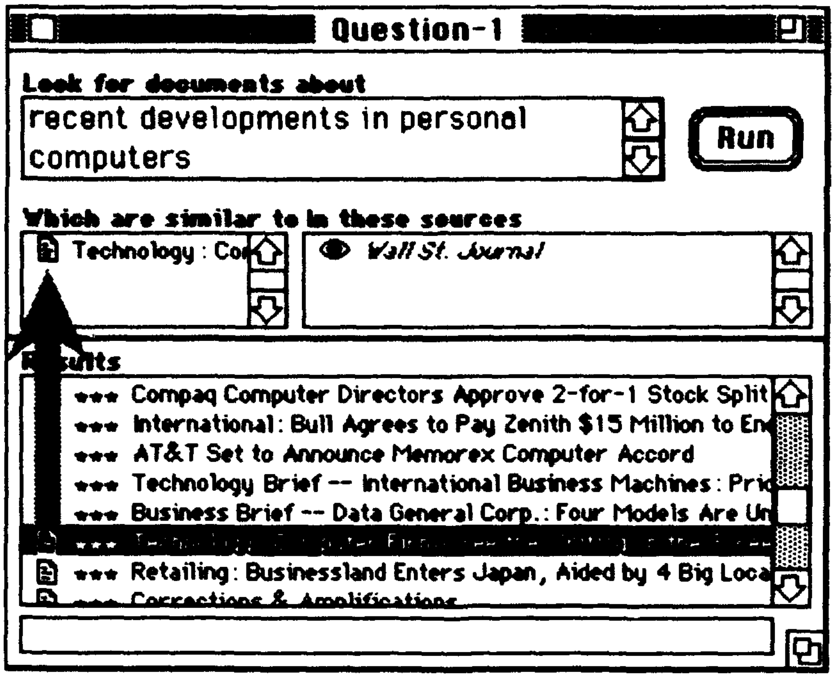


Figure 4: To refine the search, one or more of the result documents can be moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

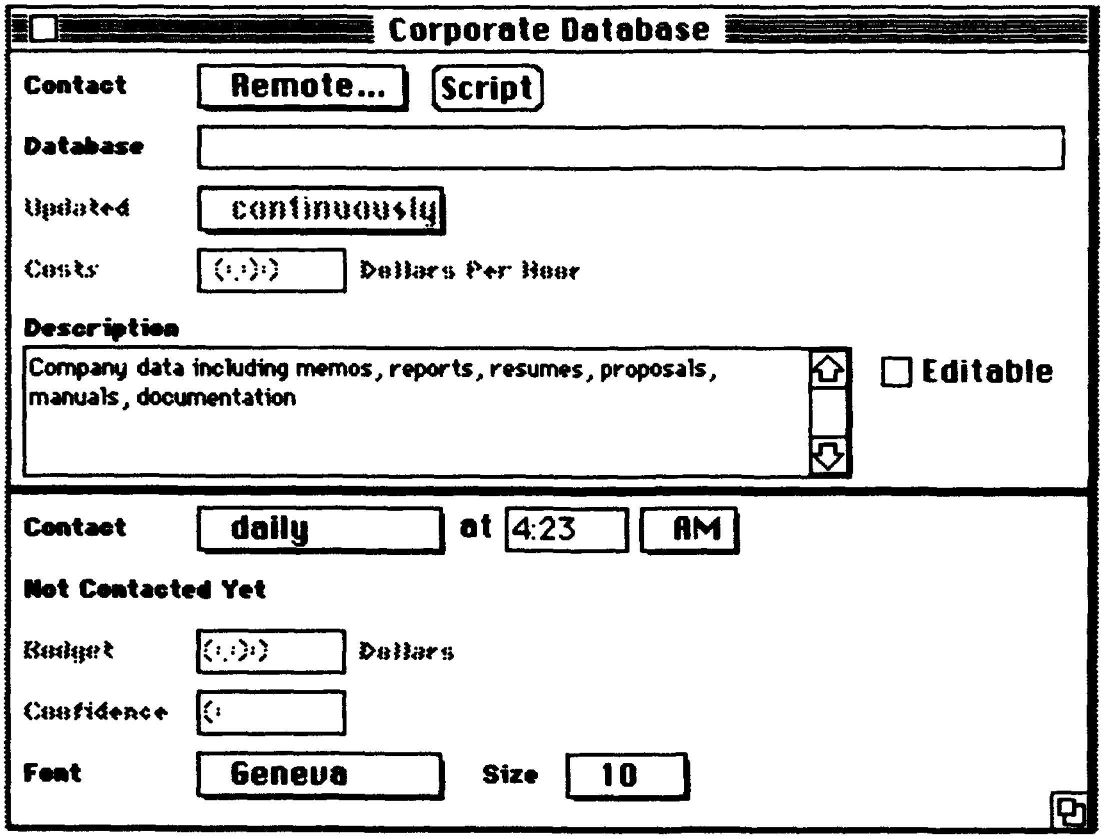


Figure 5: The Source description contains all the necessary information for contacting an information server.

The WAIS protocol is an extension of the existing Z39.50-1988 standard, but we are working with the standards committee to merge the extensions back into the newer versions (Davis, 1990). The extensions allow support for multi-media data, large documents, a directory of servers, different communication systems, and distributed retrieval. To support multi-media, a document must be available in a number of formats. This was accomplished by listing the set of available types in the search response from which the client can choose one to retrieve. Another problem with the protocol involved retrieving large records. Large documents, whether text or not, would be slow to display if the whole document had be retrieved at one time, as is required in the original standard. Large documents are supported in the WAIS protocol by allowing the client to retrieve sections of a document based on bytes or lines. We also standardized a format for describing servers (Kahle, 1991a) and how to contact them, which is necessary to implement a directory of servers. To support communication systems other than the full OSI protocol stack, a header was needed to show how long the packet was and how it was encoded. With this packet header we implemented the WAIS protocol over modems, TCP/IP, and X.25 systems. To support distributed retrieval we needed a document identifier system that could be used in a distributed environment (Kahle, 1991b).

The protocol used in the WAIS system has proven useful in the distributed full-text environments in which we tested it.

2.3 *User Interfaces: Asking Questions*

Users interact with the WAIS system through the Question interface. Each question form has an area for the user's quasi-natural language question, the list of sources which will be accessed to try to answer the query, the list of relevant documents, and a list of answer documents.

The illustrations here are taken from the initial WAIStation program produced at Thinking Machines for the Apple Macintosh. We have also built clients for X windows and gnu-emacs. Another Macintosh interface was developed that emphasizes the alerting feature (Erickson, 1991).

With most current retrieval systems, complications develop when one begins dealing with more than one source of information. For example, one contacts the first source, asks it for information on some topic, contacts the next source, asks it the same questions (most likely using a different query language, a different style of interface, a different system of billing), contacts the next source, and so on. One of the primary goals behind the development of the WAIS system was to replace all this with a single interface.

With WAIS, the user selects a set of sources to query for information, and then formulates a question. When the user presses the RUN button (see figure 2), the system automatically asks all the desired servers for the required information with no further interaction necessary by the user. Thus, the documents returned are sorted and consolidated in a single place, to be manipulated by the user. The user has transparent access to a multitude of local and remote databases.

From the user's point of view, a server is a source of information. It can be located anywhere: on the local machine, on a network, or on the other side of a modem. The user's workstation keeps track of a variety of information about each server. The public information about a server includes how to contact it, a description of the contents, and the cost. In addition, individual users maintain their own private information about the servers they use.

Users may need to budget the money they are willing to spend on information from particular servers, know how often and when each server is contacted, and assess the relative usefulness of each server. In the current interface, the budget entries were put in as place holders, since all servers are currently free. When a source is contacted, all questions that refer to the source are updated with the new results.

A "confidence factor" allowed users to multiply the score returned from different servers so that the list presented to the user would be more appropriate. This was put in the interface to anticipate a number of different server technologies with different scoring algorithms. The "confidence factor" allows the user to adjust the scores. Also, a user might have a preference for the information from one server over another so a subjective balance would be helpful. This feature was rarely, if ever, used since the number of servers was small, they all used the same server technology, and most users only asked one source at a time anyway.

2.4 *Servers*

The servers in the WAIS system hold databases that can be queried by a client. References of documents that best match the words and phrases in the query are returned to the client. A client can then request all or part of a document from the server. Since the client explicitly contacts the server, any number of billing methods could be employed such as 900 numbers, credit cards, and subscriptions.

The Connection Machine server system (CMDRS), used in the WAIS system, stores the documents in a compressed form, called signatures, which can be searched quickly using the parallel processors of the Connection Machine (Stanfill, 1986). The signatures are stored in the RAM of the machine thereby assigning a few documents to each processor of the machine. Each word in the query is then broadcast to all the processors, and a score is kept for each document to reflect the number of words and phrases that matched. Weighting is done based on crude proximity and occurrence frequency. The resulting search results have been found to be useful to end-users.

As the dissemination of information becomes easier, questions of ownership, copyright, and theft of data must be addressed. These issues confront the entire information processing field, and are particularly acute here. The WAIS system is designed to keep control of the data in the hands of the servers. A server can choose to whom and when the data should be given. Documents are distributed with an explicit copyright disposition in their internal format. This is not to say that theft cannot occur, but if a client starts to resell another's data, standard copyright laws can be invoked. By keeping the control of the distribution of works with the creators, many of the problems of copyright do not arise.

2.5 *Rerunning Questions - A Personal Newspaper*

In addition to providing interactive access to information, the WAIS system can also be used as a rudimentary personal newspaper to alert its user when new documents are available on a subject that might be of interest (see figure 6). In the library literature, this is referred to as selective dissemination of information (SDI), and many manual, semi-automated, and automated systems have been implemented. Our initial implementation is to save interactive questions and

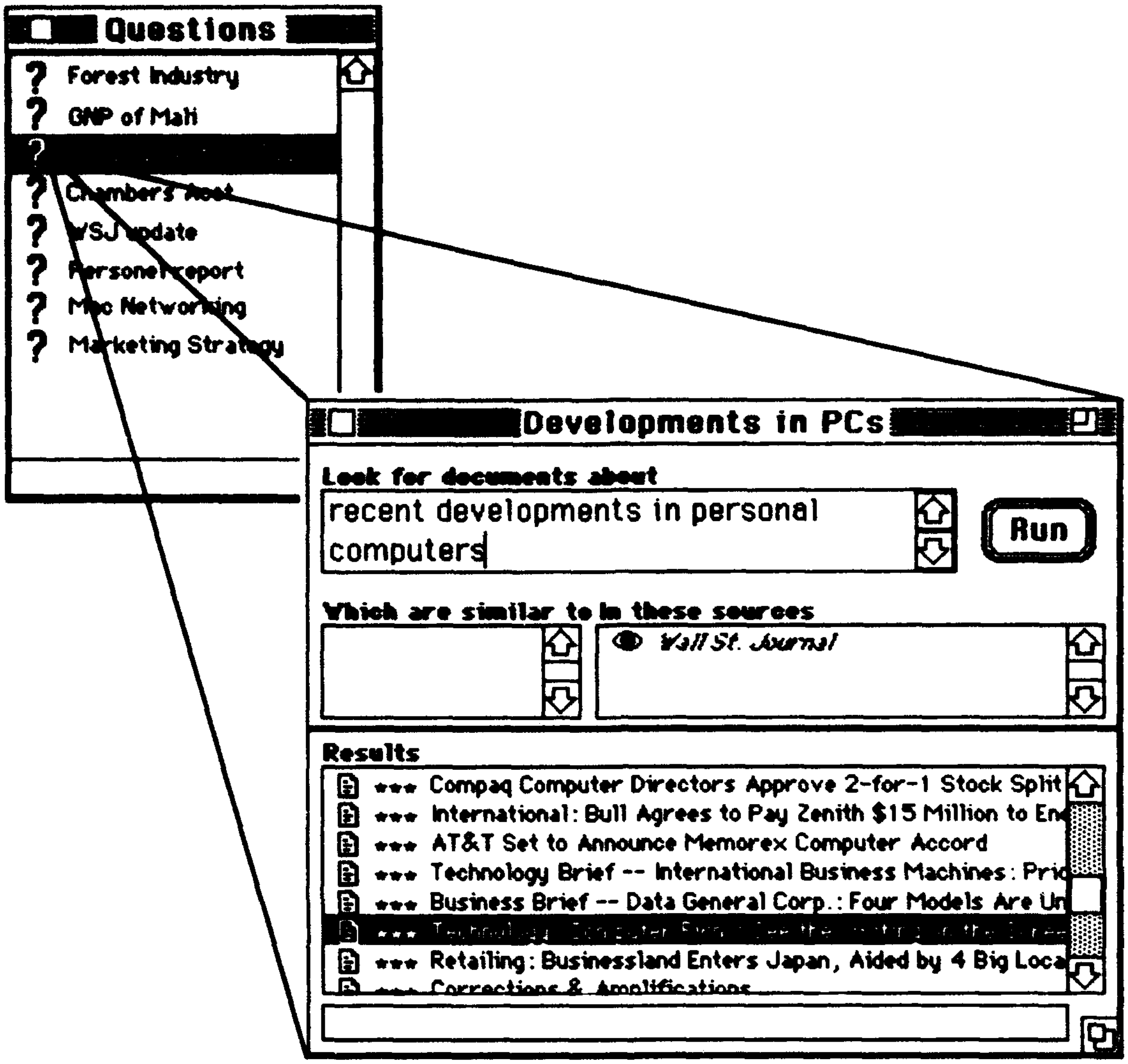


Figure 6: Opening a saved question which was automatically updated in the background, and contains new data.

automatically rerun them at periodic intervals checking if new documents were available. This technique has the advantage of hiding communication costs, using systems off hours, and finding potentially interesting information in a timely manner.

2.6 *MultiMedia Database*

The documents retrieved through WAIS may be any kind of file such as text, still graphics, motion pictures, or hypertext documents. The searching of the system is based on an initial quasi-natural language question and further relevance indications, but the server is free to use that information in any way to find appropriate documents. The protocol simply defines a document as a block of data and a type. The type is used by the client to determine how to display the document. A list of available types is part of the search response of each document. This allows clients to choose among a selection of types, and suppress documents whose types they can't display. Alternatively, they can simply store the documents in their local disk for latter processing. Our initial X windows clients are able to use other programs to display graphic data such as TIFF and GIF. The Macintosh client can display PICT images and text, but can theoretically download any type of file.

Non-textual data is indexed in one of two ways. If the data includes an embedded description (e.g. TIFF), the description is used for indexing. Otherwise an external description is indexed. When a search identifies the description file as a suitable response, the multi-media data is returned instead of the description file.

2.7 *The Directory of Servers*

To find sources of information in a distributed environment, we used a "directory of servers" which is a database of documents describing other servers. In response to a query, the database of servers is searched, returning a list of documents (i.e., server descriptions) which match the query. Instead of text documents, however, it takes advantage of the mixed type capabilities of WAIS to return a structured document with many specific fields for cost and contact information (see figure 5). This capability will become more important as the number of servers increases.

For example, suppose you needed information concerning the current gross national product of Mali, but had no idea on which server to find it. You could first ask the directory of servers for "information about the current economic condition of Mali." The directory will take the words in the query and find descriptions of the servers that contain those words. It might then return several documents. The World Factbook, for instance, might appear because of a match on "economic condition". This source description could then be used as the source field of another question. This time, the system would contact the World Factbook, ask for the information, and possibly return a document with a description of Mali (World Factbook, 1974).

Additionally, the directory of servers provides a means for information providers to advertise the availability of their data. When a new source becomes available, the developers can submit a textual description, along with the necessary information for contacting the server. This information is added to the directory, and becomes available to the public by the searching interface.

3. The Prototype WAIS System

In the fall of 1990 we installed an experimental WAIS system at Peat Marwick. The prototype was used by 20 users in six cities. Peat Marwick utilized corporate data in Montvale, New Jersey, and Dow Jones information in Princeton, New Jersey. The system was run successfully for six months with good user reactions.

KPMG Peat Marwick is an example of an information-intensive company. Their role as consultants requires that they maintain an awareness of new products, market fluctuations, changing laws, internal regulations, and competition. In addition, as a large organization, there is considerable internal information that can be leveraged, such as company contacts, bids, reports, and resumes. Furthermore, their distribution in 40 countries, with 200 offices in the United States alone, makes them a prime candidate for wide area information technology.

The primary users were located in San Jose and connected by 56kbaud and 9.6kbaud circuits to the servers in New Jersey. The 20 managers and partners in the Peat Marwick's accounting division used an 8192 processor Connection Machine system for serving reports, proposals, resumes, contracts, accounting manuals, the Peat Marwick Audit Manual, Management Guide, and Professional Development Courses, documents from the Financial Accounting Standards Board, the Government Accounting Standards Board, and the American Institute of Certified Public Accountants, and a tax library. The data were separated into twelve different databases, which could be searched separately or in any combination. There was also a virtual database consisting of all of these sources.

The connection to Dow Jones provided access to 1 Gigabyte of data, running on a 32K processor Connection Machine. The data consisted of a year of the Wall Street Journal, Barrons, and 400 magazines. Each of the approximately 250,000 articles was a separate document. The ability to search personal data was not available at the time of the experiment.

3.1 *Lessons Learned*

The search technology performed well in finding useful data for end users who were given little instruction about system use. The speed of the searches (usually between two and ten seconds) depended on the communication speed, since the search itself took much less than a second. When the response time was greater than 10 seconds, the users voiced complaints, but in general they were very pleased with the search results. The ability to execute searches without prior training, and without in-depth knowledge of the database was essential to the users. Relevance feedback was used frequently and effectively by users who were aware of its existence. Not all users realized it was available, however. This is an opportunity for improving user interfaces. For example, relevance feedback could be performed automatically on any document which the user chooses to view. This would result in a kind of automatic, dynamically linked hypertext system, where every document is "linked" to all similar documents.

The Macintosh user interface (WAIStation) also performed well in terms of ease of use and adaptability. With a single demonstration, most users were able to execute searches and save their results. Left with only the manual, new users took 15 to 30 minutes to feel comfortable with the system. The ability to transparently search local and remote databases was greatly appreciated, as reported in user feedback forms. The biggest problem we had with the interface was in implementing the TCP and modem connections from the Macintosh. The automatic updating feature of WAIStation was rarely used and needs more work to make it more obvious and to allow it to give better feedback when documents are found.

Wide area communications proved to be a difficult part of the project due to our resistance, based on future cost projections, to use leased lines. The original plan called for linking San Jose and Montvale with Shiva Telebridges™ running at 9600 baud on a normal phone line. This approach did not prove reliable, nor did it give us reasonable performance. We ended up replacing this link with a dedicated 56kbaud line attached to a SyncRouter (Engage Communications™). The dedicated line was highly reliable and 56kbaud was fast enough to support many active users of the system, while maintaining an interactive feel in both search and retrieval.

Organizing and formatting the data for display on the client workstation proved to require more effort than we expected. The current Macintosh client is capable of displaying only ascii text and PICT format picture files. This meant that the corporate data, which consisted primarily of word processor files, had to be converted to ascii. Since the conversion was not perfect, some documents required a small amount of manual reformatting. This is obviously unacceptable in a production system. A more attractive solution might be to build a client which can display the most common document formats, and which can call on other applications to display formats it doesn't understand. This approach will become easier to implement as document filters (e.g., Claris™ XTND) and interprocess communication become more common. This approach will also make it possible to index and store the original document, rather than an ascii shadow.

As the searchable Peat Marwick corporate collection grew, the users wanted to search just parts of the database. The natural divisions for the users were the original sources of the text, such as training manuals or government legal texts.

In summary, we found that the users were pleased with the system, and some used it many times each day. It appears that there is a market for end user search systems, and that the technology is ready. The weak link seems to be communication infrastructure.

4. Conclusion

In developing the WAIS system, the participating companies have demonstrated that current hardware technology can be used effectively to provide sophisticated information retrieval services to novice end users. How this might affect information providers is not yet understood. The users at Peat Marwick found the technology useful for day-to-day tasks such as researching potential new accounts and finding resources within their own organization. Since these tasks are not restricted to the accounting and management consulting industries, we are optimistic that this type of technology can be fruitful and productive in many corporate settings.

The future of this system, and others like it, depends upon finding appropriate niches in the electronic publishing domain. Potential uses include making current online services more easily accessible to end-users, and allowing large corporations to access their own internal data more effectively. It is also possible that near-term development will focus on a single professional field such as patent law or medical research.

Acknowledgments

The design and development of the WAIS Project has been a collective effort, with contributions and ideas coming from many people. Among them:

Apple Computer: Charlie Bedard, David Casseras, Steve Cisler, Ruth Ridder, Eric Roth, John Thompson-Rohrlich, Kevin Tiene, Gitta Soloman, Oliver Steele, Janet Vratny-Watts.

Dow Jones News/Retrieval: Rod Wang, Roland Laird.

KPMG Peat Marwick: Chris Arbogast, Mark Malone, Tom McDonough.

Thinking Machines: Dan Aronson, Patrick Bray, Jonathan Goldman, Danny Hillis, Rob Jones, Barbara Lincoln, Gordon Linoff, Chris Madsen, Gary Rancourt, Sandy Raymond, Steve Schwartz, Tracy Shen, Craig Stanfill, Robert Thau, Ephraim Vishniac, David Waltz, Uri Wilensky.

References

(Davis, 1990) F. Davis, B. Kahle, H. Morris, J. Salem, T. Shen, R. Wang, J. Sui, and M. Grinbaum, "WAIS Interface Protocol Prototype Functional Specification," Thinking Machines, April 1990. Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com or wais server wais-docs.src.

(Egan, 1989) D. Egan, J. Remde, L. Gomez, T. Landauer, J. Eberhardt, and C. Lochbaum, "Formative Design-Evaluation of SuperBook," ACM Transactions on Office Information Systems, January 1989.

(Erickson, 1991) T. Erickson and G. Salomon, "Designing a Desktop Information System: Observations and Issues," Proceedings of the ACM Human Computer Interaction Conference, 1991. ACM Press, April 1991, pp. 49-54.

(Ginther-Webster, 1990) K. Ginther-Webster, "Project Mercury," AI Magazine, 1990, pp. 25-26.

(Kahle, 1989) B. Kahle, "Wide Area Information Servers Concepts," Thinking Machines technical report TMC-202, November 1989. Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com or wais server wais-docs.src.

(Kahle, 1991a) B. Kahle and H. Morris, "Source Description Structures," February 1991. Available via anonymous ftp: /pub/wais/doc/source.txt@quake.think.com.

(Kahle, 1991b) B. Kahle and H. Morris, "Document Identifiers or International Standard Book Numbers for the Electronic Age," May 1991. Available via anonymous ftp: /pub/wais/doc/doc-ids.txt@quake.think.com.

(Malone, 1986) T. Malone, K. Grant, and F. Turback, "The Information Lens: An Intelligent System for Information Sharing in Organizations; In Human Factors In Computing Systems," CHI'86 Conference Proceedings (Boston, MA): ACM: New York, 1986; pp 1-8.

(NISO, 1988) "Z39.50-1988: Information Retrieval Service Definition and Protocol Specification for Library Applications," National Information Standards Organization (Z39), P.O. Box 1056, Bethesda, MD 20817, (301) 975-2814. Available from Document Center, Belmont, CA. Telephone 415-591-7600.

(Salton, 1971) G. Salton, "The Smart Retrieval System - Experiments in Automatic Document Processing," Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1971.

(Salton, 1983) G. Salton and M. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, New York, 1983.

(Stanfill, 1986) C. Stanfill and B. Kahle, "Parallel Free-Text Search on the Connection Machine System," Communications of the ACM 29,12, December 1986, pp.1229-1239.

(Stanfill, 1991) C. Stanfill, "Massively Parallel Information Retrieval for Wide Area Information Servers," International Conference on Systems, Man, and Cybernetics, Charlottesville, Virginia, October 1991.

(World Factbook, 1974) Publishing Sciences Group, Acton, Massachusetts, 1974.

INFORMATION RETRIEVAL SERVICES ON WIDE AREA NETWORKS: A VISION OF THE FUTURE

by

J. LeMézec
Vice President, Science & Technology
France Telecom North America
1270 Avenue of the Americas
Suite 2800
New York, NY 10020
United States

Summary

This paper will refer to some existing systems and to R&D work, especially in France Telecom, in order to help define a vision of future systems and services that store and retrieve information on Wide Area Networks
Special reference will be made to the Minitel network, a videotex network involving 6 million terminals and 20,000 services; JPEG "Minitel Photo" and the first Fast Speed Teletel services have been experimented in 1992. Other services, prepared for short-term experimentation, include multimedia videotex on ISDN.
Two main contributions to a future broadband multimedia network will then be discussed: standardization of multimedia and the deployment of ATM (Asynchronous Transfer Mode).
Finally, a vision of a multi-purpose worldwide network of networks will be proposed and the challenges risen by its construction will be presented.

1. Introduction

The databases where information is stored and the networks on which they can be accessed are as diverse as the needs and interests of their potential users. It is highly desirable that any person can access this information worldwide, whatever is the specific network he or she is connected to. Before addressing this challenge, we shall present some comments on existing networks which are the starting points to build the networks on which information should be retrieved in the future.

2. About some existing information networks and their short-term evolution.

This part of the paper will focus on Videotex and on ISDN networks, especially on those which have been developed by France Telecom. However I shall first make a few remarks on the Internet, although briefly, since it is described and discussed in other papers during this conference.

2.1. The Internet and the public data networks

The Internet is a network of networks. It is growing at a tremendous rate both in its US birthplace and internationally. It is also becoming increasingly commercial. Boosted by the US Presidential program HPCC (High Performance Computing and Communications) and as a result of its innate growth, the Internet is experiencing a dramatic (r)evolution and it is at the heart of the "Information Superhighway" and of the "National Information Infrastructure" debates in the United States.
However, this growth has also given rise to some critical problems, such as the "address crisis" (the shortage of addresses in the Internet Protocol) which could necessitate in-depth modifications in this protocol or even its replacement by something else.
Nonetheless, and whatever the future will be, we must acknowledge the outstanding contribution of the Internet community to the creation of the existing information networks and services. Besides the development of the TCP/IP protocol itself, it is relevant to mention such services as the Internet E-mail, FTP, GOPHER and, more recently, WAIS and MIME. These services are or will be available worldwide.

As a consequence of this success, many people consider that the information network of the future will result from successive upgrades of the Internet. The vision we propose is somewhat different. Designed to serve the scientific community, the Internet has not introduced from the beginning some of the constraints of the conventionnal public networks, including the cost of production and delivery of information, intellectual property rights issues, privacy and availability of the services: to introduce these constraints, a process now in progress, is not a small task.

Moreover, large amounts of information relevant to scientists and technicians are stored outside the Internet, in the public telecommunication networks. These networks have been developped in order to address the needs of a variety of users, and not only those of the scientific community. They are generally based on CCITT and OSI Standards rather than on TCP/IP. As an example, France Telecom, has developed, through its subsidiary Transpac, a large X.25 network recently upgraded

Presented at an AGARD Meeting on 'International High Speed Networks for Scientific and Technical Information', October 1993.

at 2 Mbit/s. However, Transpac is now offering multiprotocol access on its high bit rate network, including IP interfaces and Frame Relay. A large number of information servers connected to Transpac network can be accessed through the French PSTN (Public Switched Telephone Network), as part of the value-added Videotex service well-known under the name of Minitel.

We shall add that the French National Research and Education Network (Renater) which was built and is operated by France Telecom offers an IP service as well as an X.25 service, partially based on Transpac. Renater is connected to the NSFnet, one of the national backbones of the US Internet, through a 2 Mbit/s transatlantic link. So, the two worlds, -the world of the Internet and the world of OSI, -are communicating, in spite of some technical difficulties, and we can say that the latter extend the reach of the Internet beyond its own realm.

2.2. The French Videotex Network

We will now consider another kind of information network, taking as an exemple the French Videotex and its short term evolution.

Videotex is an interactive service defined by CCITT to allow users accessing information stored in remote data bases through a telecommunication network. Although many countries have developed a videotex service, France is the real place where it turned out to be a definite success.

This success, which would had never happened without an appropriate regulation, was partly due to the technology used, but probably much more to a deep consideration of human factor issues and to the partnership with many potential service providers. The success also required a strategy, high investment and time.

Although the commercial name of the French videotex service is TELETEL, it is commonly known as MINITEL, the commercial name of the specific terminal which has been developed for this service in France. However, it would be inexact to mix up the videotex concept with a specific terminal. In fact, TELETEL can be accessed through any Personal Computer with an emulation software. Both, a Minitel terminal or a PC can be used in North America to access remote data bases connected to TELETEL in Europe or even located in the US or in Canada. A PC with a Minitel emulation software can also give access to the Internet. More and more countries are being connected to TELETEL through a gateway called *Minitelnet*.

To give some figures, 6.3 million Minitel terminals are in use in France (2/3 in homes, 1/3 in offices), to be compared with the 30 million subscriber lines of France. There are 20,000 independent services and the revenue was 1 Billion US$ in 1992.

A small number of these services deal with scientific or technical information. As an example, the engineering guidelines for France Telecom network are recorded on a Teletel server (with a restricted access, for sure). Other Scientific and Technical Information Service Providers on Teletel are Cedocar (restricted access), Euro (Thesis and Laboratories Repertory), Gaphyor (Bibliographic Database on Physics and Physicochemistry) and others.

The architecture of this network (fig.1)is of interest here, since we will use it to sketch the future information networks. The terminals are connected to the conventional PSTN (Public Switched Telephone Network) through analog telephone lines and a modem. The servers are connected to the Public X.25 Data Network operated by Transpac. These two networks are connected through Gateways, the Videotex Access Points (PAVI: Point d'Accès Videotex). The PAVIs, communicating points between a synchronous network (the PSTN) and an asynchronous network (TRANSPAC) have the functions of a PAD (Packet Assembler and Dissassembler) and other functions such as presentation of the available services, support for the dialog between the user and the system, and billing. The intelligence of the system is located in the PAVIs and in the servers. However some of the functions of the PAVIs could be transferred to the PSTN, as the Intelligent Network architecture is implemented in the telecommunication networks.

In fact, the French videotex service is not rigid and will evolve in the future.

Direct access to Teletel services are being offered to customers who have a LAN with an X.25 output and who can be directly connected to Transpac through a Private Access Point (PAP: Point d'Accès Privé). This solution is similar to that used in the Internet where users are directly connected to a data network.

Another evolution, now in progress, is related to the commercialization of new Minitel terminals, more adapted to professional usages. For instance, the Minitel 12 is both a multifunction telephone set and an intelligent videotex terminal, with an optional vocal electronic receptionist and an optional telematic message recorder.

A third evolution, also in progress, is a high speed Teletel service (TVR: Teletel Vitesse Rapide) operating at 4.8 or 9.6 Kbit/s, in lieu of the conventional 1.2 Kbit/s. This service is open since May 26, 1992 and last year 23 clients where experimenting this service in partnership with France Telecom. It enables a photographic display according JPEG Standard.

A further step will be Videotex on ISDN (TGV: Teletel Grande Vitesse): we shall discuss this evolution in the next paragraph.

These successive speed upgrades make possible to implement on the Minitel terminals dialogs and graphical interfaces used on PC terminals (use of windows, mice, icons,...), an evoluton which would provide another step to the convergence between the Minitel world and the computer world.

2.3 Information Retrieval on ISDN Networks.

ISDN is an upgrade of the telephone network which gives to subscribers a digital access at 144 Kbit/s (2 B channels at 64 Kbit/s and a D channel at 16 Kbit/s) or 2 Mbit/s (30 B + D; in North America: 24 B + D = 1.5 Mbit/s). It is available nationwide in France, the EC Directives for an Open Network Provisioning of ISDN in Western Europe are under discussion and both the US Government and Telecommunication Network Operators have announced ISDN as a national priority. By now, international ISDN links are currently available.

ISDN offers an enhanced facility to retrieve information from remote databases (Fig.2). We shall present a few such applications right now available or under development in France.

The first one is a system commercialized by SARDE SA and resulting from a Project developed by CNET (the R&D Laboratory of France Telecom) between 1983 and 1987. SARDE stands for Système d'Archivage et de Recherche de Documentation Electronique (in English: Electronic Documentation Storage and Retrieval System). The documents (maps, diagrams, drawings,...) are digitized, compressed and stored on optical disks such as CDROMs associated to databases. They can be accessed through a 64 Kbit/s channel with the support of an Electronic File Management System. A TCP/IP gateway can be provided. More than 100 such systems are in operation. One of them has been developed in order to computerize the storage and the management of the technical documentation relevant to the 2000 switches operated by France Telecom. This documentation represents 2 million pages in constant evolution (at a rate of 15% per year). By the end of 1992, 300 Minitel and fax terminals and 55 ISDN workstations (ultimately 650) were connected to this system.

Another example is a picture mailbox on ISDN operated by IMAGE DIRECTE. Through this system. a number of photo agencies offer access in a few seconds to their collection of millions of pictures. Using their computer screen, clients can select their pictures by comparing photos from different agencies and order them.

We have already mentioned Videotex on ISDN. Existing applications are related to sales transactions, tourism agencies, entertainment. Other applications are related to distant learning, medicare, meteorology, and, of course, scientific and technical information retrieval. The important point here is that it provides a path to Broadband Multimedia Information Retrieval Services.

3. Broadband Multimedia Information Retrieval Services.

Multimedia is now a popular word, but it is used with a variety of meanings. For our purpose here, we shall say that a terminal, a PC, a system, a service or an application is multimedia whenever another media is used together with texts and graphics and whenever there are some synchronization and interactive manipulation functions between the medias involved. These functions can be provided either by the terminal, the information server or the network itself. An example of the later case is given by the multimedia communication software MEDOS developed by Nynex. In the future, these three solutions will be applied concurrently.

We shall define Broadband or High Bit Rate Information Services by the availability of at least 2 Mbit/s at the user interface. In the future, such services could imply the availability of gigabit/s at this interface.

Multimedia does not imply Broadband. Some kind of Multimedia communication can be provided through TVR (Teletel Vitesse Rapide) or even through the use of screen phones and a communication protocol. However, in order to provide enriched services supporting moving pictures or high quality displays, it is necessary to use broadband transmission lines and large storage capacities.

We shall now focus on two main topics related to the emerging broadband multimedia services.

3.1 Multimedia Standardization Issues.

The existing applications generally refer to JPEG (for still images) and to MPEG (for moving pictures) compression standards which have made possible the implementation of multimedia resources on a large number of computers. Following the same process which gave rise to these standards, a group of experts called MHEG (Multimedia and Hypermedia Experts Group) is defining an international standard which will provide "an encoding for multimedia/hypermedia information that can be used and interchanged by applications in a wide range of domains".

The MHEG standard uses an object-oriented analysis but does not require a conforming process to be object-oriented. The specification addresses the needs of minimal resource terminals and makes use of other standards for the component Text, Image, Graphic and other objects. Additionally, the MHEG Objects provide abstractions suited to real-time presentation (through multimedia synchronization functions) and real-time interchange (with minimal buffering using normal speed data communication). The standard defines classes of objects. It provides an ASN.1 representation of the Objects and an alternate SGML representation. The MHEG Group is installing at Lowell University (USA) a FTP server which will provide public domain MHEG softwares and MHEG documentation under Postscript format. An introduction to this standard and a demonstration were presented at the ACM 93 Multimedia Conference (Anaheim, USA, August 1993).

In the mean time, another group, known as The Interactive Multimedia Association (IMA) is also active in promoting multimedia technologies and applications, through a "Compatibility Project" and through the creation of Multimedia Forums. IMA's Compatibility Project is intended to provide a technical forum "for the development of recommendations for cross-platform compatibility of multimedia data and applications". Requests For Technology (RFT's) have been released: one of them deals with Multimedia Services, a second with A Scripting Language For Interactive Multimedia Titles, and a third with Multimedia Data Exchange.

Last, but not least, the Internet Community has been active in developing its own multimedia applications and standards: MIME, or Multi-purpose Internet Mail Extensions, is one of them.

So, the main issue does not reside in a lack of standards for multimedia services. On the contrary, a plethora of standard families could result into incompatibilities, giving rise to non-communicating information worlds.

3.2 ATM, a basic technique for the future Multimedia Broadband Networks.

A good news for those who are interested in intercommunicating information worlds is the growing consensus to consider ATM as a basic technique to build the broadband networks of the future. ATM (Asynchronous Transfer Mode) is a data transmission and switching technique which puts the information into fixed-length packets called cells. It combines the flexibility of conventional packet-transmission with the simplicity of ordinary digital telephone switching. Following laboratory demonstrations, it has been adopted by the CCITT as the basic technique for Broadband-ISDN (B-ISDN) Networks, intended to transmit moving pictures as well as speech and data. Resulting from a compromise between parties focused on different applications, the cell length has been fixed to 53 Bytes, including a 48 Bytes information payload and a 5 Bytes Header. All the Telecommunication Network Operators are committed to considering ATM as a target for future B-ISDN Networks.

However, the use of ATM will not be restricted to public telecommunication networks. Mainly through the activity of the ATM Forum, an organization where both representatives of the computer and the telecommunication industries as well as users meet, ATM has become a strong candidate for providing a seamless communication technique from LANs to MANs and WANs. As a result, a variety of products are being offered commercially and a number of field tests or even operational deployments are in progress. VISA, one of the US Gigagit test beds, located in North Carolina, was probably the first ATM field trial in operation. More recently, ATM has been chosen as the technology for the upgrade of the NASA Science Internet and of the Energy Science Internet, two important components of the future US NREN, and thus of the future Internet. In Europe, 17 Telecommunication Operators have signed a Memorandum of Understanding for building a pan-european network based on national ATM Cross-Connects: Requests For Provisioning are in progress and the network will partly be in operation next year. In the mean time, many European operators are deploying ATM equipments in their networks for both trials and operation. France Telecom, for instance, is preparing the introduction of ATM equipments in the French National Research and Education Network RENATER. We can be confident that in a few years there will exist a high bit rate ATM link between the US NREN and RENATER.

A more recent aspect of the introduction of ATM in the future information networks is the reference which is made to ATM for the design of Video-Dial-Tone (VDT) or Video-On-Demand (VOD) Platforms, proposed by CATV Operators as well as Telecommunication Operators. These systems are based on a high capacity storage medium, where can be stored, for instance, some thousand of movies, with a number of ATM accesses. ATM should provide flexibility and a better use of the transmission resources, resulting in cost reduction, a major factor for the development of broadband techniques. Such a video database could be replaced by a multimedia database where scientific and technical information could be stored on a media such as CD-ROMs. These two applications have in common to be interactive; both will need a high bit rate to deliver the information to the user and a much lower bit rate for the dialog between the user and the server.

ATM provides a certain level of compatibility between networks. In fact it acts at level 2 of the OSI Protocol Stack, the data link level. ATM is able to carry several kinds of traffics, but to do this some mapping is necessary. The mapping is achieved by the ATM Adaptation Layer which is known by its abbreviation AAL. CCITT has defined four classes of traffic and their adaptation layers:

-AAL1: Connection-oriented, constant bit rate, e.g. PCM voice

-AAL2: Connection-oriented with timing, variable bit rate, e.g. compressed video

-AAL3: Connection-oriented, variable bit rate, e.g. data service (bursty traffic)

-AAL4: Connectionless, variable bit rate, e.g. SMDS (in Europe, CBDS)

Furthermore, an AAL5 has been developped in order to provide "a simpler and more efficient AAL" than the AAL3/4 for LAN interconnection.

This multiplicity of AALs may result on one hand in some compatibility issues or at least some complexity. On the other hand, every kind of traffic may be carried and mixed which is of interest for multimedia services. Moreover, the discussions about connection-oriented versus connectionless networks which gave rise to religious type debates become of second tier interest. As a matter of fact, successful tests have been conducted in order to demonstrate IP traffic transmission on long distance ATM links, a result which is important for the future of the Internet.

However such a common basis for the future communication networks, although facilitating their connexion, is not sufficient to ensure a complete interoperability, which involve also the higher levels of the protocol stacks. The future networks should also coexist and interoperate with much lower bit rate networks which will be in operation and probably will continue to grow for a number of years.

4. The coexistence of low-speed and very-high-speed information services.

Low-speed data services may seem outdated now that gigabit-per-second are on the horizon. However, for many years, a majority of users of telecommunication networks will not have sophisticated terminals at their disposal. More small- and medium-sized companies, schools and social organizations as well as residential users (sometimes for home working or for learning or entertainment) will be entering the information market, but cannot afford expensive terminals. Moreover, some specific applications will need low-speed access. For instance people traveling with their laptop will need access to information through a low-speed dial-up modem. According to some experts, these new usages could contribute to the rebirth of low-speed data services.

From our point of view, the more important consequence is the co-existence of a wide range of access speeds within the future information networks, from kilobit- to gigabit-per-second. This parameter should be taken into account when designing and implementing the information servers. The information could be organized in layers corresponding to different grades of information. For instance, a first level could handle a very simple kind of information, such as text which could be delivered at low speed to simple terminals. A second level could deal with more complex information, including sound and still pictures, delivered to ISDN or PC terminals at a medium speed (TVR: Teleservice à Vitesse Rapide) or even at a higher speed (TGV: Teleservice à Grande Vitesse). A third level could handle huge amounts of information (complete books, video documents or movies for instance) stored on Interactive CDs, and delivered to workstations through "Information Superhighways", probably based on ATM, through a communication protocol which has still to be invented. One can also imagine that the server or the data network could recognize the class of the terminal and its access characteristics, in order to determine the level of information which may be delivered and the type of dialog which should be used.

5. The vision of a worldwide multi-purpose information network.

The US Government has announced, as one of his primary goal, "the construction of an advanced National Information Infrastructure (NII), a seamless web of communications networks, computers, databases, and consumer electronics that will put vast amounts of information at users' fingertips," and his will to "promote seamless, interactive, user-driven operation of the NII".

At a time when we are entering in the Electronic Information Age, this objective could be proposed as a worldwide goal, which will certainly be welcome by AGARD Technical Information Panel Members.

This future infrastructure will be a "network of networks". In each country these networks will be different in some aspects. The diversity of present-day networks from which the future networks will emerge, the differences in regulations and economic development are factors resulting into local variations and even into different fundamental choices. Moreover, in a majority of countries, different networks will co-exist.

The vision of the infrastructure we are proposing here, is derived from Videotex networks (Fig.3). It relies on a large number of information servers, some of them multimedia, perhaps some of them organized according to the design suggested in the preceding paragraph, in order to address the requirements of different grades of users and terminals. These servers will be connected to high-speed digital networks, generally based on the ATM technique, often at gigabits-per-second speed: we shall call these networks Data Transport Networks, even though they will transport sound, still and moving pictures as well as data. The data transmission will be performed on a number of SDH (Synchronous Digital Hierarchy, SONET in North America) channels optically multiplexed according to a WDM (Wavelength Domain Multiplexing) technique and providing up to terabit-per-second (i.e. millions of Mbit/s) capacity. The terminals will generally be connected to an access network, which could be narrowband, and which will perform the main function of transmitting the customer's requirements to the server. The Access Network will be connected to the Data Network through a Gateway. The information will be delivered to the customer through the Broadband Data Transport Network and a Broadband Distribution Network. Gateways will also interconnect different Access Networks or different Data Transport Networks, resulting in a network of networks.

From this basic architecture, we can imagine different variations which will coexist in the future. First, some Information Databases could be directly connected to the Distribution Network, as it is the case for some Video Databases which will be connected to the Video-Dial-Tone Platforms. Second, some customers will be connected to the main Access Networks through specific communication links. This will be the case for PCS or other mobile terminals, or for users having at their disposal complex Customer Premise Equipments (CPE) such as LANs or PBXs. Moreover some large information consumers will have a direct access to the Data Transport Network.

The intelligence needed to operate and maintain such a network of networks, including functions such as the allocation of the information and communication resources, the supervision of the quality of the delivered service, the checking of access rights, the billing of users, will be distributed between the Servers, the Data Transport Networks, the Access Networks, the Gateways and the Customer Premises including the Terminals. Some of these Intelligence islets will be interconnected as it is already the case for the existing Intelligent PSTN, to constitute an Overlay Control Network.

5. Conclusion: the challenges of a vast undertaking.

After presenting some remarks on existing information retrieval networks and their short-term evolution, we have proposed a vision of a multi-speed, multi-purpose information infrastructure.

Building this worldwide network of networks is a vast, multibillion-dollar undertaking. Fortunately, it will not have to be built from scratch: many components exist and are in operation; some others will result from an evolution of existing networks which is already under progress. However, in order to reduce the cost, this infrastructure should be, at least partly, multipurpose, serving a wide range of users, and not just a specific category of users. Using a common infrastructure should reduce the cost charged to each customer.

The interoperability of the heterogeneous basic networks involved is the main technical issue. Fortunately the modem techniques can solve such problems. However, standardization would simplify the solutions and result in a more efficient and cost-effective infrastructure. In addition, the use of wide-spread equipments should also result in dramatic cost reductions.

However, cost and technical problems are not the only challenges to be faced. Some other issues are the following:
-Privacy, security of the information and network reliability;
-Protection of intellectual property rights;
-Protection against the blocking of the network bottlenecks resulting from overwhelming traffic and unnecessary information;
-Navigation through this worldwide network to locate and retrieve the required information;
-More generally, the definition of user-friendly interfaces to this complex information world.

As usual, the solution of these problems will require to strike a balance between regulation and freedom, and between competition and cooperation. It will mandate the active participation of all concerned parties, including the Internet Community,, Telecommunication Network Operators and Information Service Providers.

Acknowledgements

I should like to acknowledge my colleagues of France Telecom, specially those from our R&D Centers (CNET, -Centre National.d'Etudes des Télécommunications, -and its subsidiary CCETT, -Centre Commun d'Etudes de Télécommunication et de Télédiffusion) and those from France Telecom North America and Minitel USA, who helped me to prepare this review, either by giving access to their personal documentation or by useful discussions. However, this paper reflects my personal vision of the information network of the future, at the time it has been written, in a world characterized by rapid changes in regulations, market and technology, which make difficult any prevision of the future.

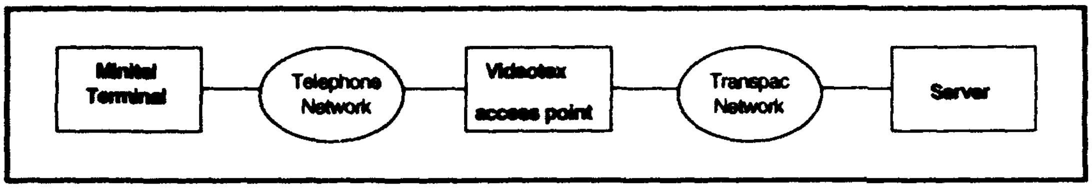


Figure 1: The value-added Videotex Network for TELETEL

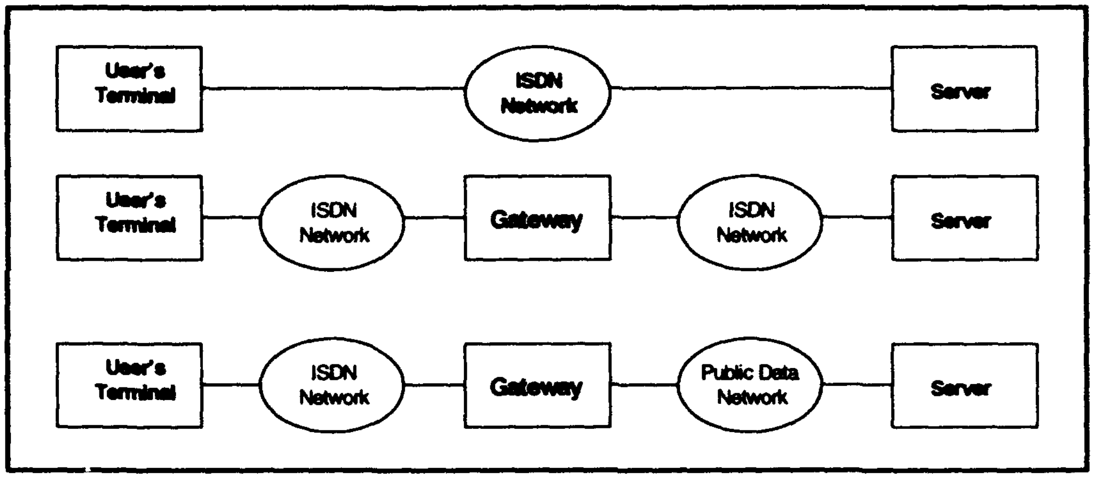


Figure 2 : Three Configurations for retrieving *data, image or multimedia information into* ISDN *terminals.*

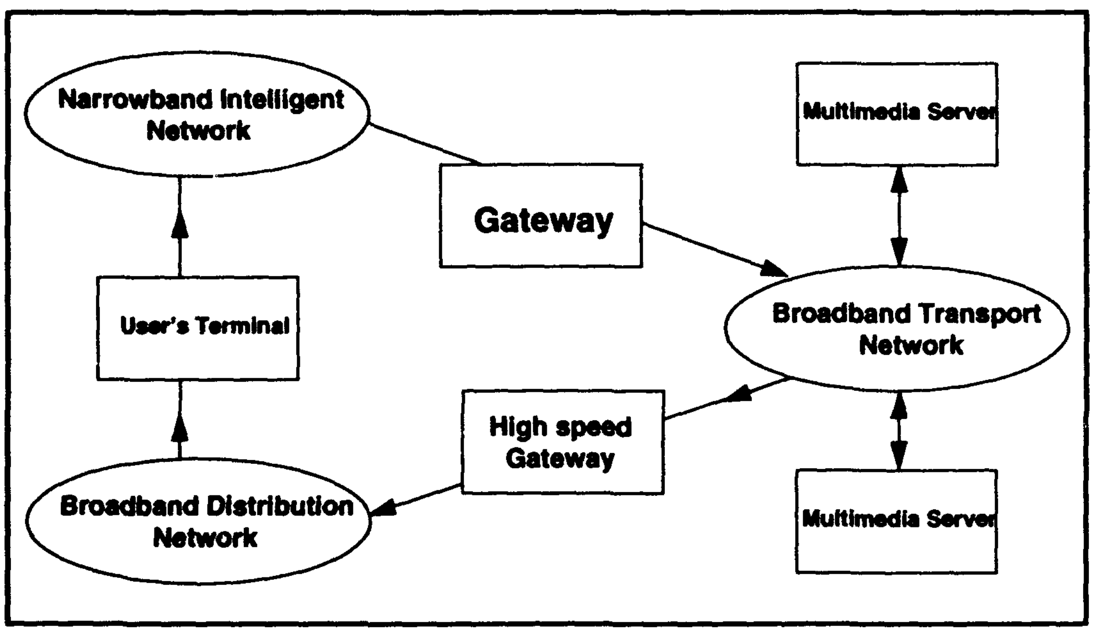


Figure 3 : Basic architecture of a future broadband information network

INTERNATIONAL HIGH SPEED NETWORKS FOR SCIENTIFIC AND TECHNICAL INFORMATION

Technical Information Panel Specialists' Meeting
Ottawa, Canada 6th—7th October, 1993

LIST OF PARTICIPANTS

Ms K. ADAMS	Canadian Library Assn, 602-200 Elgin Street, Ottawa, Ontario K2P 1L5, Canada
Ms M. AMBA	Micromedia Ltd, 20 Victoria Street, Toronto, Ontario M5C 2N8, Canada
Mrs E. ARBACH	Vanier Library, University of Ottawa, 11 Marie Curie Street, Ottawa, Ontario K1N 9A4, Canada
Mr R.M. ARMSTRONG	Defence Research Establishment, Atlantic (DREA), c/o FMO Halifax, Halifax, N.S. B3K ZXO, Canada
Ms B.F. AUBREY	Head, Document Delivery, DSIS, National Defence Headquarters, MGeneral George R. Pearkes Building, Ottawa, Ontario K1A 0K2, Canada
Ms J. BAKER	DSIS 2-2 Information Scientist, National Defence Headquarters, Ottawa, Ontario K1A 0K2, Canada
Mr C.J. BIGGER*	Chief Librarian, GEC Marconi Research Centre, West Hanningfield Road, Great Baddow, Chelmsford, Essex CM2 8HN, United Kingdom
Mr W. BLADOS	c/o NASA Scientific and Technical Information Program, NASA Headquarters, (Code JTT), Washington DC 20546, United States
Mr A. BODINI†	ESA/ESRIN, Projects & Engineering Dept, Telecommunications Unit, Via Galileo Galilei, 00044 Frascati, Italy
Ms H.T. BOOTH	Chief Librarian, National Energy Board, 311-6th Ave. S.W., Calgary, Alberta T2P 3H2, Canada
Mr M. BRANDRETH†	Chief, Policy, Planning & Systems, Canada Institute for Scientific and Technical Information, National Research Council of Canada, Ottawa, Ontario K1A 0S2, Canada
Mrs D. BRAUN	National Research Council, CISTI, Uplands Branch, U-61, Ottawa, Ontario K1A 0R6, Canada
Mr G. BRILLET	CNET, LAA/TTP/RIC, 22301 Lannion, France
Mr J.L. CADIEUX	Chief, Library Services, Transport Canada (AFMHG), 15th Floor, Tower 'C', 330 Sparks Street, Ottawa, Ontario K1A 0N5, Canada
Dr S. CAKIR*	Director, TUBITAK-EDB, Informatics Division, Scientific & Technical Research Council of Turkey, Ataturk Bulvari No. 221, 06100 Kavaklidere, Ankara, Turkey
Ms J.M. CAMERON	CSE Library, PO Box 9703, Terminal, Ottawa, Ontario, Canada
Ms D. CANNING	Computer Services Librarian, McGill University, 809 Shersbrooke Street, West Montreal, Quebec H3A 2K6, Canada
Ms S. CARNSON	Science Librarian, Science Library, A&AEE Boscombe Down, Salisbury, Wilts SP4 0JF, United Kingdom
Mr V. CASTELO†	Centro Tecnico de Informatica del CSIC, Calle Pinar 19, 28006 Madrid, Spain
Ms L. CHAILLEZ	CRDV — Bibliotheque, C.P.8800, Courcelette, Quebec G0A 1R0, Canada
Dr P.B. CHARLESWORTH	Head, Computer Technology, Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8, Canada
Mr G. CLEVELAND†	National Library of Canada, Information Technology Services, 395 Wellington Street, Ottawa, Ontario K1A 0N4, Canada
Mr J.A. COLVIN	Micromedia Ltd, 20 Victoria Street, Toronto, Ontario M5C 2N8, Canada

† Author
* Panel Member

Mrs M.F. COMBEROUSSE	Ministere de l'Enseignement Superieur et de la Recherche/SDIST, 1 rue Descartes, 75231 Paris Cedex 05, France
Ms M. CUTLER	Canadian Agriculture Library, 930 Carling Avenue, Ottawa, Ontario K1A 0C5, Canada
Mr E. DANIEL	Canadian Agriculture Library, 930 Carling Avenue, Ottawa, Ontario K1A 0C5, Canada
Dr A. DEL REY*	Head U.E.I. Information Retrieval, (CSIC) C.I.N.D.O.C., c/o Joaquin Costa 22, 28002 Madrid, Spain
Dr M. DYKSTRA	School of Library & Information Studies, Dalhousie University, Halifax, NS B3H 4H8, Canada
Ms T. EGAN	Information Specialist, US Department of Energy, DP/23, Washington DC 20585, United States
Mr T. FOTHERINGHAM	Transport Canada (AFCHC), Tower 'A' Place de Ville, Ottawa, Ontario K1A 0N5, Canada
Mr A.D. GALE	McLaughlin Library, University of Guelph, Guelph, Ontario N1G 2W1, Canada
Ms L. GAMBLE	Health & Welfare Canada, H.P.B., Library Services Division, Banting Bldg, Ottawa, Ontario K1A 0L2, Canada
Mr A. GARDEAU†	Software Development, CAE Electronics Limited, PO Box 1800, Saint Laurent, Quebec H4L 4X4, Canada
Mr J.A. GILBERT	John A. Gilbert and Associates, Telecommunication Consultants, 714 Highland Avenue, Ottawa, Ontario K2A 2K7, Canada
Dir. R.A. GJERSVIK*	The Technical University Library of Norway, The University of Trondheim, Hogskoleringen 1, N-7034 Trondheim, Norway
Mr N. GOLDMANN	ARRAY Development Inc., PO Box 5145 Stn. F, Ottawa, Ontario K2C 3H3, Canada
Ms J. GOODMAN	Canadian Agriculture Library, 930 Carling Avenue, Ottawa, Ontario K1A 0C5, Canada
Mr W.F.M. GREGOIRE	International HQ and Support Command, Staff Support Branch, B-7010 SHAPE, Belgium
Dr M.C. GUTIERREZ* (Deputy Panel Chairman)	Div. de Materiales y Estructuras, INTA, Crrtra. Torrejon a Aljalvir, Km.4, 28850 Torrejon de Ardoz, Madrid, Spain
Dr H. HAEGE*	Streitkraefteamt/abt.III, FIZBw, Friedrich-Ebert-Allee 34, 53113 Bonn, Germany
Dr D.B. HEARN	Defence Research Agency, St Andrews Road, Malvern, Worcestershire WR14 3PS, United Kingdom
Mr T.A. HERMANN*	Head, Customer Service & Marketing, CASI, Center for Aerospace Information, Baltimore MD 21240-0757, United States
Mr R. HEROUX	Departmental Library, Canadian Heritage, 15 Eddy Street, Hull, Quebec K1A 0M5, Canada
Mr M. HILLER	Ecole Polytechnique de Montreal, C.P. 6079, Succursale A, Montreal, Quebec H3C 3A7, Canada
Ir P.J. HOOGENBERK*	Head of Scientific & Technical Documentation Centre of the Netherlands Armed Forces (TDCK), PO Box 90701, 2509 LS The Hague, The Netherlands
Ms J. HUNTER†	NASA Scientific and Technical Information Program, 1213 Jefferson Davis Highway, Suite 1200, Arlington, Virginia 22202, United States
Mr P. IMHOF	Head, Technical Information Division, Naval Research Laboratory, Washington DC 20375, United States
Mr D. JOHNSON	Reference & Information Services, National Library of Canada, 395 Wellington Street, Ottawa, Ontario K1A 0N4, Canada

† Author
* Panel Member

Ms C. JOLING	Centre Library, IDRC, Box 8500, Ottawa, Ontario K1G 3H9, Canada
Mr B. KAHLE†	President, Wide Area Information Servers, WAIS Inc., 1040 Noel Drive, Ste 102, Menlo Park, CA 94025, United States
Mrs R. KAMRA	6136 Voyageur Drive, Orleans, Ontario K1C 2W3, Canada
Prof. P.T. KIRSTEIN	University College London, Gower Street, London WC1E 6BT, United Kingdom
Mrs J. KRALIK	Library Manager, Industry & Science Canada, 235 Queen Street — 03 West, Ottawa, Ontario K1A 0H5, Canada
Mr E.T.H. LAPEYSEN*	Wetenschappelijk raadgever, NCWTD — Koninklijke Bibliotheek, Keizerslaan 4, B-1000 Brussels, Belgium
Mr J-C. Le MOAL	Unité de Communication et de l'Information Scientifique INRIA, BP 105, 78153 Le Chesnay Cedex, France
Mr J. LeMEZEC†	Vice President Science & Technology, France Telecom North America, 1270 Avenue of the Americas, Suite 2800, NY, New York 10020, United States
Ms J LEONARDO	National Research Council, CISTI — Parkin Branch, Montreal Road, Bldg M-2, Ottawa, Ontario K1A 0R6, Canada
Ir H.A.M. LUIIJF	TNO-FEL, PO Box 96864, 2509 JG Den Haag, The Netherlands
Mr C. LUPOVICI†	Directeur de la Production, INIST, 2 Allée du Parc de Brabois, 54514 Vandoeuvres-les-Nancy Cedex, France
Ms C. MacKEIGAN†	Canada Institute for Scientific and Technical Information, National Research Council of Canada, Ottawa, Ontario K1A 0S2, Canada
Ms C. MacLENNAN	DND/DSIS, Box 291 BBW RR 3 Woodlawn, 4421 Tranquility Lane, Ottawa, Ontario K0A 3M0, Canada
Mr J. MARKHAM	Director of Marketing — Europe, MPR Teltech Ltd, 2000-320 Queen Street, Ottawa, Ontario K1R 5A3, Canada
Ms M. McCONNELL	Health & Welfare Canada, H.P.B., Library Services Division, Banting Bldg, Ottawa, Ontario K1A 0L2, Canada
Ms S.G. McINTYRE	Defence Research Establishment, Ottawa, Ontario K1A 0Z4, Canada
Dr C. McKNIGHT†	HUSAT Research Institute, The Elms, Elms Grove, Loughborough, Leicestershire LE11 1RG, United Kingdom
Ms P.E. MEINDL	AECL Research, Chalkriver Laboratories, Chalkriver, Ontario K0J 1J0, Canada
Mr M. MEYER	Centre d'Information/Documentaire, Centre Commun de Recherches Louis-Bleriot, 12 rue Pasteur — BP 76, 92152 Suresnes Cedex, France
Ms M. MONTGOMERY	Director General, Canada Institute for Scientific & Technical Information, Montreal Road, Bldg M-55, Ottawa, Ontario K1A 0S2, Canada
Ms T. NARAYNSINGH	Project Manager, Library Systems, Supreme Court Building, Library, Wellington Street, Ottawa, Ontario K1A 0J1, Canada
Ms A. ÖZDIL*	Ministry of Defence (MSB), Department of R&D (ARGE), 06650 Bakanliklar-Ankara, Turkey
Mr W.J. PADDEN†	Industry and Science Canada, Information Technologies, Industries Branch, 235 Queen Street, Ottawa, Ontario K1A 0H5, Canada
Mr O. PARADIS	Ecole Polytechnique de Montreal, Bibliotheque, C.P. 6079, Suce. A, Montreal, Quebec H3C 3A7, Canada
Ms M. PATRIARCKI	DND/Library, 101 Colonel By Drive, Pearkes Library — 3 North Tower, Ottawa, Ontario K1R 7T8, Canada
Mrs D. PATRINOU*	Technical Scientific Library, Hellenic Aerospace Industry, PO Box 23, GR 32009 Schimatari, Viotias, Greece

† Author
* Panel Member

Mr B. PAWLOWSKI	Pratt & Whitney Canada Inc., 1000 Marie Victorin, Longueuil, Quebec, Canada
Ms M. PERRIER	Chief, Public Services, Library, Communications Canada, 300 Blater, 14th floor, Ottawa, Ontario K1N 0C8, Canada
Ms H. POWELL	Library, Technical University of Nova Scotia, PO Box 1000, Halifax, Nova Scotia B3J 2X4, Canada
Ms D. RABOW	Canadian Agriculture Library, 930 Carling Avenue, Ottawa, Ontario K1A 0C5, Canada
Mrs C. RENAUD-FRIGON	Communications Canada, 300 rue Slater, Ottawa, Ontario K1A 0C8, Canada
Lt Col. J.G. RENY (Rtd)*	Senior Administrator, Military Agency for Standardization, NATO/MAS, B-1110 Brussels, Belgium
Ms R. REPSYS	DND/DSIS-4, MGen G.R. Pearkes Bldg, National Defence Headquarters, Ottawa, Ontario K1A 0K2, Canada
Mrs V. RITCHIE	Library Manager, Forestry Canada, 351 St Joseph Blvd, 17th Floor, Hull, Quebec K1A 1G5, Canada
Mr D. ROBINSON	National Library of Canada, Room 215, 395 Wellington Street, Ottawa, Ontario K1A 0N4, Canada
Dr J.L. ROBINSON	SHAPE Technical Centre, PO Box 174, 2501 CD The Hague, The Netherlands
Mr S.J. ROBY	Telesat Canada, 1601 Telesat Court, Gloucester, Ontario K1B 5P4, Canada
Capt. R.A.G.B. ROQUE*	CDIFA, Base de Alfragide, Av. Leite de Vasconcelos, 2700 Amadora, Portugal
Mr J. ROSS	Associate Director, National Technology Transfer Center, Wheeling Jesuit College, Duvall Center, 200 Mt De Chantal, Wheeling, WV 26003-6563, United States
Mr R.P. RYAN*	Deputy Administrator, Defense Technical Information Center, Cameron Station, Alexandra, VA 22304-6145, United States
Prof. J-M. SALAUN	Dir. Etudes et Recherches, ENSSIB, 17-21 Bd du 11 novembre 1918, 69623 Villeurbanne Cedex, France
Mr M.J. SCHRYER*	Director, Directorate of Scientific Information Services, National Defence Headquarters, MGeneral George R. Pearkes Bldg, Ottawa, Ontario K1A 0K2, Canada
Mrs D. SCHRYER	41 Pheasant Run Dr., Nepean, Ontario K2J 2R3, Canada
Dr M. SCOTT	The National Librarian, The National Library of Canada, 395 Wellington Street, Ottawa, Ontario K1A 0N4, Canada
Mr R. SEARLE*	Chief Librarian, Ministry of Defence, 3-5 Gt Scotland Yard, London SW1A 2HW, United Kingdom
Mr R.W. SLANEY	Head of Library, MOD Broadoaks, Parvis Road, West Byfleet, Surrey KT14 6LY, United Kingdom
Mr A.W. SMITH	Head, Special Acquisitions, The British Library Document Supply Centre, Boston Spa, Wetherby, West Yorks LS23 7B9, United Kingdom
Mr G.A. STEPHENSON†	Knowledge Exchange Technologies Ltd, 11 rue du Plebiscite, L-8037 Strassen, Luxembourg
Mrs J. STITT	Canadian Agriculture Library, 930 Carling Avenue, Ottawa, Ontario K1A 0C5, Canada
Ir A.S.T. TAN*	Information Specialist, National Aerospace Laboratory (NLR), PO Box 90502, 1006 BM Amsterdam, The Netherlands
Mr R. TEXLER	NASA Lewis Research Center, Mail Stop 60-1, 21000 Brookpark Road, Cleveland, Ohio 44135, United States
Mr F. TSIMPOGLOU*	Nat. Hellenic Research Foundation, Nat. Documentation Center, 48 Vas. Constantinou Ave, GR-116 35 Athens, Greece
Mr E. VALENSI*	Directeur, C.E.D.O.C.A.R., 00460 Armées, France

† Author
* Panel Member

Ms M. VanBUSKIRK†	Canada Institute for Scientific and Technical Information, National Research Council of Canada, Ottawa, Ontario K1A 0S2, Canada
Mr J. VILLAIN*	Chef du Dépt Information/Strategie, Société Européenne de Propulsion, 24 rue Salomon-de-Rothschild, BP 303, 92156 Suresnes Cedex, France
Ms C. WALKER*	Head, Information Services, SHAPE Technical Centre, PO Box 174, 2501 CD The Hague, The Netherlands
Dr M.H. WALSHAW*	Head, Defence Research Information Centre, MOD, Room 2160, Kentigern House, 65 Brown Street, Glasgow G2 8EX, United Kingdom
Ms T. WHALEN	ORA Library, National Defence Headquarters, 101 Colonel By Drive, Ottawa, Ontario K1A 0K2, Canada
Mr W. WHALEN	DL/DOA, 2690 C Street, Suite 4, Wright Patterson AFB, Ohio 45433-7411, United States
Ms C. WILTSHIRE	Micromedia Ltd, 20 Victoria Street, Toronto, Ontario M5C 2N8, Canada
Mr M. WOLFE	Director of Marketing, Faxon/SMS, PO Box 5087, STNF, Ottawa, Ontario K2C 3H3, Canada

† Author
* Panel Member

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-544	ISBN 92-835-0736-3	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 Rue Ancelle, 92200 Neuilly sur Seine, France

6. Title
INTERNATIONAL HIGH SPEED NETWORKS FOR SCIENTIFIC AND TECHNICAL INFORMATION

7. Presented at the Technical Information Panel Specialists' Meeting held at the National Library of Canada, Ottawa, Canada, 6th—7th October 1993.

8. Author(s)/Editor(s)	9. Date
Various	February 1994

10. Author's/Editor's Address	11. Pages
Various	104

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Information systems
Information technology
Computer Networks
Communication networks
Wide area networks
Electronic mail
Internet — network

14. Abstract

Contains the papers presented at the Technical Information Panel's Specialists' Meeting held in Ottawa, Canada, 6th—7th October 1993.

Subjects covered include descriptions of Internet and other global information networks, the virtual library, the supply of documents, electronic journals, bulletin boards, electronic mail and conferencing, information systems for unstructured files, standards, training, and a view of the future.

END